

UNIT 6

Graphs

Introduction

This unit extends your work on graphs, which you began in Unit 2, by looking at some connections between algebra and graphs. You have seen that relationships between quantities can often be expressed using algebraic formulas. For example, in Unit 2 you saw a formula expressing the relationship between the distance, speed and time for a journey, and a formula expressing the relationship between the number of days for which you hire a car and the cost of hiring it. This unit is about relationships between *two* quantities, such as the car hire relationship, or the relationship between distance and time if you travel at a constant speed. Relationships between two quantities can be visualised by using graphs, and this unit concentrates on relationships that correspond to *straight lines* on graphs. You will see other types of relationships between two quantities later in the module.

Section 1 shows you how to use a formula expressing the relationship between two quantities to draw a graph of the relationship. It also shows you how sets of paired data can sometimes be modelled by straight lines on graphs.

Section 2 concentrates on two key characteristics of a straight-line graph: its slope (how steep it is) and its position. You will learn how to measure the slope of a line, and how to interpret it. The slope of a straight-line graph indicates how one quantity changes in comparison with the other, and it can give you valuable information about the relationship between the two quantities that would be difficult to spot from a table of numbers.

In Section 3 you will see that every straight-line graph can be described by an algebraic equation and learn how this algebraic equation is linked to the characteristics of the graph. In mathematics it is often useful to be able to look at a problem both algebraically and graphically, so the ideas in this section are important and you will need them in future units. In this section you will also meet a particular type of algebraic relationship between quantities, *direct proportion*, which occurs in many practical situations. The graphs that illustrate relationships of this type are all straight lines that pass through the origin.

In Section 4 you will apply what you have learned in earlier sections to some real-world problems in which data are modelled by straight lines on graphs. In particular, you will see models for the growth of the world's tallest-ever man and investigate some possible relationships in the backache data that you met in Unit 4.

I Plotting graphs

This section shows you how to plot a graph to illustrate an equation or formula involving two variables. (Remember that a formula is just an equation that has a *subject* – a variable that appears by itself on one side of the equals sign and not at all on the other side.) The first subsection provides some brief revision of graphs and coordinates, and in the second subsection you will see an example of an equation plotted as a graph, and have the opportunity to plot some graphs yourself. In the third subsection you will see how to plot a *scatterplot* to illustrate paired data, and why this can be useful.

The idea of a *subject* of a formula was introduced in Unit 2, Subsection 3.1.

1.1 Graph axes and coordinates

The graph axes form an important part of every graph. In general discussions about graphs, it is standard to label the horizontal and vertical axes with the letters x and y , respectively, and the axes are then referred to as the x -axis and the y -axis. Each axis should have scale markings and arrows, as shown in Figure 1, and the distance between two consecutive integers on an axis is referred to as 1 unit. The graph axes are often extended to include negative numbers, as illustrated.

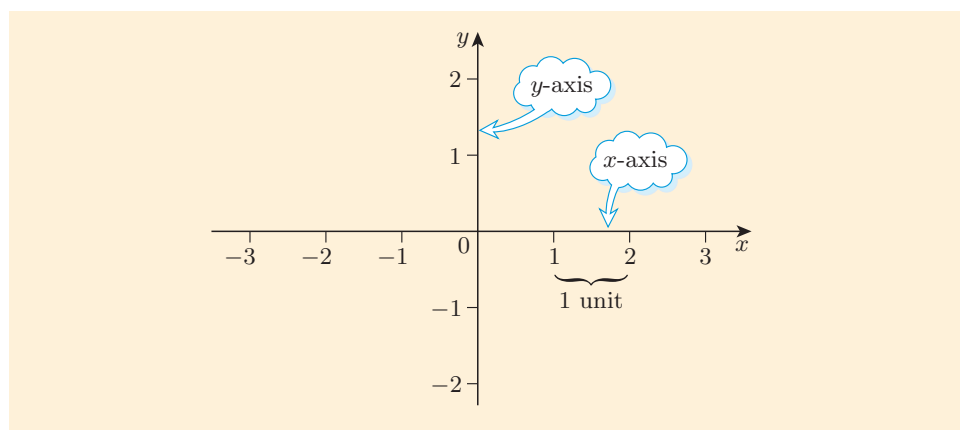


Figure 1 Standard graph axes

As you saw in Unit 2, each point that can be plotted on a graph is represented by a pair of numbers called the **coordinates** of the point. The first number specifies the position of the point along the x -axis from 0, and the second number specifies its position along the y -axis from 0. These two numbers are called the x - and y -coordinates of the point, respectively. Figure 2 shows the positions of some points, and their coordinates.

The French mathematician René Descartes (see page 37) developed this way of specifying the position of a point in 1637. It is known as the **Cartesian coordinate system**, after him.

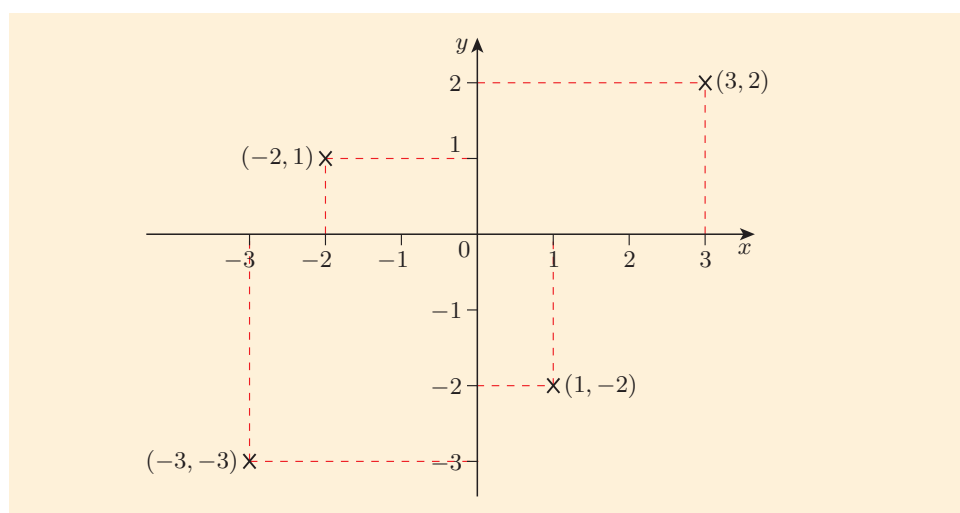


Figure 2 The coordinates of some points

Each point that is plotted on a graph can be labelled with its coordinates, as in Figure 2, or with a letter, or with both. For example, the point labelled $(3, 2)$ in Figure 2 could alternatively be labelled P or $P(3, 2)$.

If you are drawing a graph by hand, it can be useful to use graph paper so that you can plot points and read off their values accurately. However, for many purposes it is sufficient to use a blank sheet of paper, and measure

Recall from Unit 2 that you can use either a small cross or a dot to mark a point on a graph.

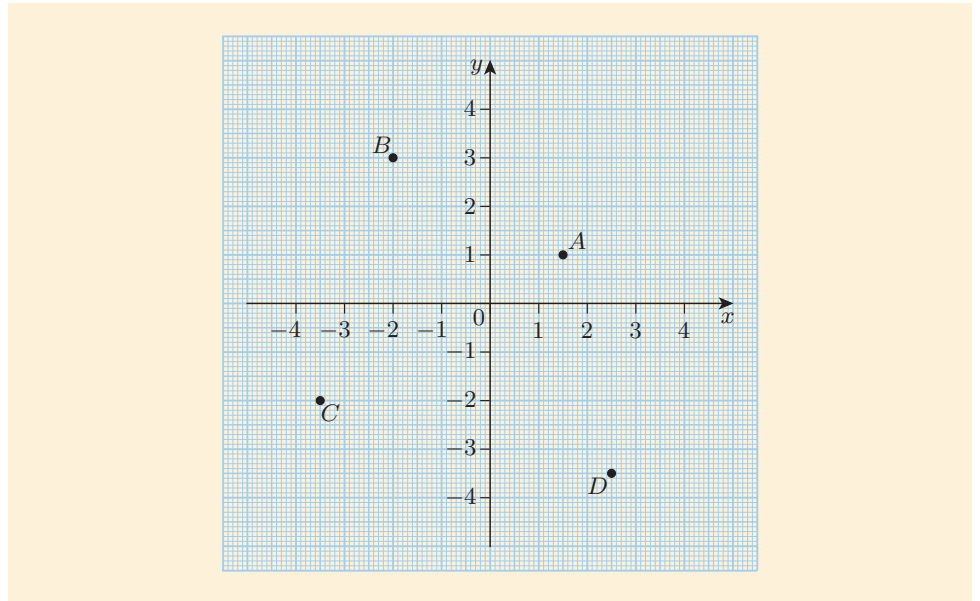
the positions of the points using a ruler or set square with millimetre markings. You can also use square-ruled paper. There are examples of graphs with all three types of background throughout this unit.

You can print out graph paper and square-ruled paper from the module website.

Activity I Writing down coordinates and plotting points

For help with coordinates, see Maths Help Module 5, Section 3.

- (a) Write down the coordinates of the points A , B , C and D shown below.



- (b) Mark the following points on a graph. You can either mark them on the graph in part (a), or draw your own axes and mark the scales.
- (i) $E(-1, -3)$ (ii) $F(0.5, 2.75)$
 - (iii) $G(-3.5, 3.5)$ (iv) $H(4, -1)$

Although x and y are the standard labels for graph axes, other labels can be used. For example, if you are drawing a graph to illustrate the relationship between the variables c and f , then you should use these letters to label the axes. Alternatively, you can use the names (and units, if appropriate) of the quantities that these variables represent.

The words used for a graph are adjusted according to the axis labels. If the axes are labelled c and f , then they are called the c -axis and the f -axis, and a point on the graph has a c -coordinate and an f -coordinate. You can also refer to the horizontal and vertical axes, and horizontal and vertical coordinates.

No matter how the axes are labelled, the first number in a pair of coordinates always gives the position along the *horizontal* axis, and the second number always gives the position along the *vertical* axis.

1.2 Graphs of equations

You can plot a graph to illustrate an equation that relates the variables x and y by first choosing some values of x and working out the corresponding values of y . This will give you the coordinates of some

points to plot on the graph. It is convenient to record the values of x and y in a table of values, as illustrated in the example below.

Example 1 Plotting a graph to illustrate an equation

Plot a graph to illustrate the equation

$$y = \frac{1}{2}x + 3.$$

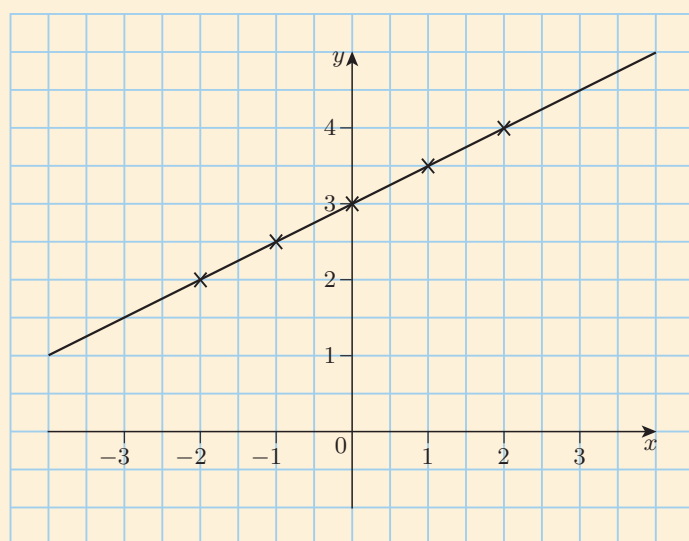
Solution

Construct a table of values. Choose some equally-spaced numbers for x , and work out the corresponding values of y by substituting into the equation. For example, substituting $x = -2$ into the equation gives $y = \frac{1}{2} \times (-2) + 3 = -1 + 3 = 2$.

A table of values for the equation $y = \frac{1}{2}x + 3$ is as follows.

x	-2	-1	0	1	2
y	2	2.5	3	3.5	4

Draw the axes and plot the points. They seem to lie in a straight line, so draw the straight line through them.



In Example 1 a straight line was drawn through the five points plotted. This is because if you were to choose some more values of x , find the corresponding values of y and plot these new points on the graph, then you would find that they all lie on the same straight line. Although this has not been checked for the example above, later in the unit you will see how you can be sure of this just by looking at the form of the equation relating x and y .

So the points that lie on the line in Example 1 are the points that have coordinates (x, y) such that

$$y = \frac{1}{2}x + 3.$$

Because of this, we say that $y = \frac{1}{2}x + 3$ is the **equation** of the line, and we often refer to the line as 'the line $y = \frac{1}{2}x + 3$ '. Only part of the line is shown on the graph – it actually extends infinitely far in each direction.



You can check whether any particular point lies on the line $y = \frac{1}{2}x + 3$ by checking whether the coordinates of the point **satisfy** the equation of the line. In other words, you have to check whether the equation is correct when the coordinates are substituted in. A convenient method of doing this for an equation like $y = \frac{1}{2}x + 3$ is illustrated in the next example.

Example 2 Checking whether coordinates satisfy an equation

Determine whether each of the following points lies on the line $y = \frac{1}{2}x + 3$.

- (a) $(8, 7)$ (b) $(-8, -3)$

Solution

 For each pair of coordinates, substitute the x -coordinate into the equation and check whether you obtain the y -coordinate. 

- (a) Substituting $x = 8$ into the equation gives

$$y = \frac{1}{2} \times 8 + 3 = 4 + 3 = 7.$$

So the point $(8, 7)$ lies on the line.

- (b) Substituting $x = -8$ into the equation gives

$$y = \frac{1}{2} \times (-8) + 3 = -4 + 3 = -1.$$

So the point $(-8, -3)$ does not lie on the line.

Activity 2 Checking whether coordinates satisfy an equation

Determine whether each of the following points lies on the line $y = \frac{1}{2}x + 3$.

- (a) $(6, 5)$ (b) $(-5, 0.5)$

When you plot a graph to illustrate an equation, the word ‘graph’ can be used to refer either to the whole picture, including the axes and all the other elements, or just to the line or curve on the picture that illustrates the equation. So the line plotted in Example 1 is the **graph** of the equation $y = \frac{1}{2}x + 3$.

In the next activity you are asked to plot the graph of another equation.

Activity 3 Plotting a graph to illustrate an equation

Complete the table of values below for the equation

$$y = -3x - 1,$$

and hence plot a graph to illustrate the equation.

x	-2	-1	0	1	2
y					

In the next activity, you are asked to plot a graph to illustrate a practical formula in which the variables are letters other than x and y . In a formula like this, the subject is known as the **dependent variable**, since its value depends on the value of the other variable, and the other variable is known as the **independent variable**.

When you plot a graph of a formula like this, you should put the independent variable on the horizontal axis, and the dependent variable on the vertical axis. In other words, you should plot the dependent variable **against** the independent variable.

In the table of values, the independent and dependent variables should be in the first and second lines, respectively, so that the coordinates will be in the correct order when you read down the columns.

Activity 4 Plotting the graph of a practical formula

You saw in Unit 2 that the formula for converting a temperature from Celsius to Fahrenheit is

$$f = 1.8c + 32,$$

where f is the temperature in degrees Fahrenheit and c is the temperature in degrees Celsius.

- (a) Which variable is the independent variable?
- (b) Construct a table of values for the formula, using the values -20 , 0 , 20 and 40 for the independent variable.
- (c) Plot a graph to illustrate the formula, drawing a straight line through the points plotted. (Use graph paper if you have some available.)
- (d) Use your graph to make the following conversions.
 - (i) Convert 10°C to $^{\circ}\text{F}$.
 - (ii) Convert -10°C to $^{\circ}\text{F}$.
 - (iii) Convert -10°F to $^{\circ}\text{C}$.
 - (iv) Convert 0°F to $^{\circ}\text{C}$.

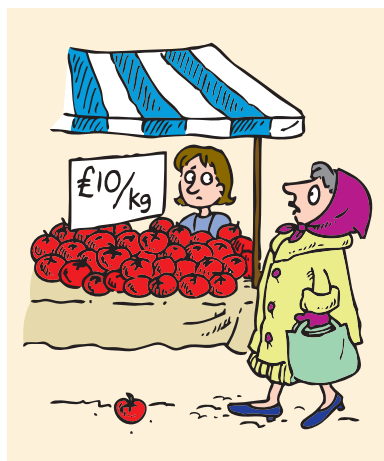
All the relationships that you have seen in this subsection have graphs that are straight lines. Relationships of this type are called **linear relationships**. You will meet other types of relationship later in the module.

The method that you have seen for plotting the graph of a relationship – constructing a table of values and plotting points – can be used to plot many types of relationship.

1.3 Scatterplots

In the previous subsection you saw how to plot the graph of an equation that expresses the relationship between two quantities. Sometimes when you are dealing with two quantities, you don't know an equation relating them – you just have some paired data that give you information about how the quantities are related. As you saw in Unit 2, in this situation you can plot the paired data on a graph to give you a visual idea of the relationship between the quantities. When you do this you might find that the points plotted lie at least approximately in a straight line.

For example, Table 1 presents some data collected by a greengrocer. She recorded the price that she charged for tomatoes, and the quantity of tomatoes that she sold, for a number of weeks. The greengrocer was



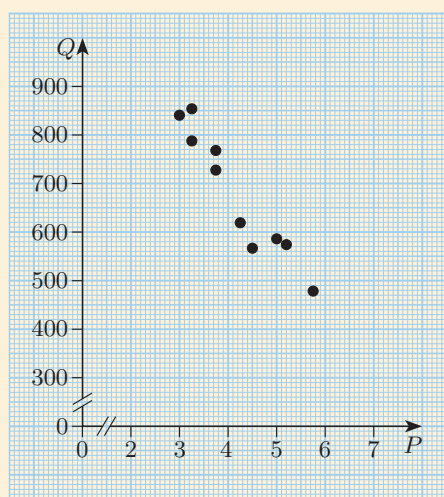
interested in the relationship between these two quantities because every week she decides what price she will charge for tomatoes (this depends on the price that she has to pay for them at the market) and then she needs to obtain the quantity of tomatoes that she is likely to sell at that price.

In the table the price in £ is denoted by P and the quantity sold in kg is denoted by Q .

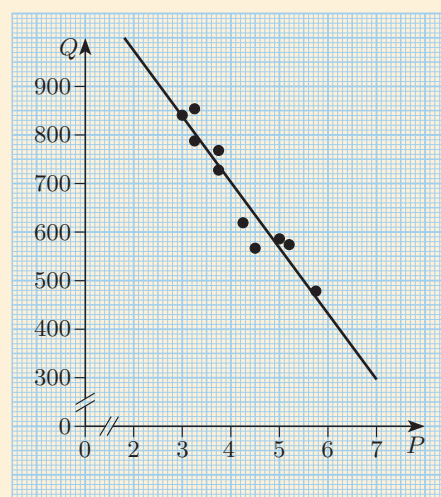
Table 1 P , price in £ charged for tomatoes, and Q , quantity in kg sold

P	3.00	3.25	3.25	3.75	3.75	4.25	4.50	5.00	5.20	5.75
Q	841	787	852	728	769	618	568	587	574	479

To produce a graph of the data pairs in Table 1, you can put price on the horizontal axis and quantity sold on the vertical axis, and plot the points (3, 841), (3.25, 787), (3.25, 852), and so on. The graph is shown in Figure 3(a).



(a)



(b)

Figure 3 (a) Q , quantity in kg of tomatoes sold, plotted against P , price in £. (b) The same graph with a straight line superimposed.

Remember that the pairs of angled parallel lines in the axes in Figure 3 indicate breaks in the scales.

A graph on which data pairs are plotted is called a **scatterplot**, and the points plotted are referred to as **data points**. The data points on the scatterplot in Figure 3(a) do not lie exactly in a straight line, so the relationship between the selling price of the tomatoes and the quantity sold does not correspond exactly to a straight line. However, the data points lie approximately in a straight line, so the relationship can be *modelled* by a straight line. A suitable line has been added to the scatterplot in Figure 3(b). Lines like this can be used to obtain useful approximate answers.

Activity 5 Using the greengrocer's graph to make predictions

For each of the following selling prices, use the graph in Figure 3(b) to predict the approximate quantity of tomatoes that the greengrocer will sell at that price.

- (a) £3.50 per kg (b) £4.75 per kg

From your work in the previous subsection, you might expect that there is an equation in the variables P and Q whose graph is the line in

Figure 3(b). There is indeed such an equation, and it would be helpful for the greengrocer to know what it is, because then she could use it to work out the quantity of tomatoes that she is likely to sell, rather than having to read it off the graph. In Section 2 of this unit you will learn about some characteristics of straight lines on graphs, and in Section 3 you will see how to use these characteristics to find the equation corresponding to any straight line on a graph.

The straight line in Figure 3(b) seems to model the data points reasonably well, but it is not clear whether it is the *best* straight line to model them. In the final section of the unit, you will see a method for finding the best straight line in this sort of situation, and a measure of how well such a line fits the data points.

2 Characteristics of straight-line graphs

In Section 1, and also in Unit 2, you have seen that a graph that illustrates the relationship between two quantities can be a straight line. This section is about the main features of graphs of this type.

2.1 Gradient

One of the main characteristics of a straight line is how steep it is. The steepness of a line is measured by its *gradient*. This subsection explains how to calculate the gradient of a line, and later in the section you will see how gradients are interpreted in practical situations.

To understand what is meant by the gradient of a line, imagine tracing the tip of your pencil along the line. The **gradient** (or *slope*) of the line is the number of units that your pencil tip moves up as it moves 1 unit horizontally to the right. For example, Figure 4 shows the line that passes through the points $(1, 2)$ and $(3, 8)$. You can see from the grid on the graph that for each 1 unit your pencil tip moves to the right, it moves up by 3 units. So the gradient of this line is 3.

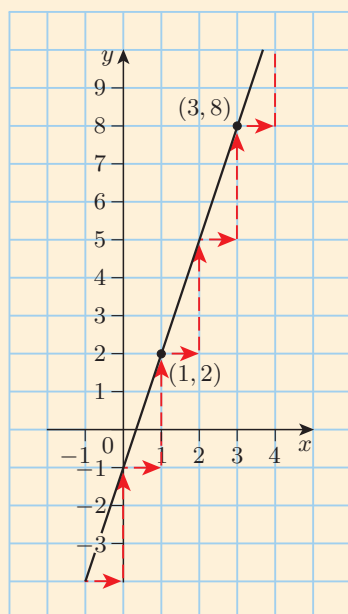


Figure 4 The line through $(1, 2)$ and $(3, 8)$

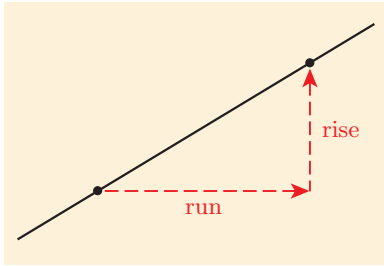


Figure 5 The run and rise

Often the gradient of a line is not as obvious as in Figure 4, and it may not be a whole number: for example, your pencil tip might move up 2.37 units as it moves one unit to the right. One way to work out the gradient of a line is to choose any two points on the line and consider what happens as your pencil tip moves from the left-hand point to the right-hand point. The increase in the x -coordinate is known as the **run**, and the increase in the y -coordinate is known as the **rise**, as illustrated in Figure 5. The gradient of the line can be calculated by dividing the rise by the run.

To illustrate this method, consider again the line in Figure 4, which passes through the points $(1, 2)$ and $(3, 8)$. As your pencil tip moves from the left-hand point $(1, 2)$ to the right-hand point $(3, 8)$, the x -coordinate increases from 1 to 3, which is an increase of 2, and the y -coordinate increases from 2 to 8, which is an increase of 6. So the run is 2 and the rise is 6, as shown in Figure 6. Therefore the gradient is $6/2 = 3$, as expected.

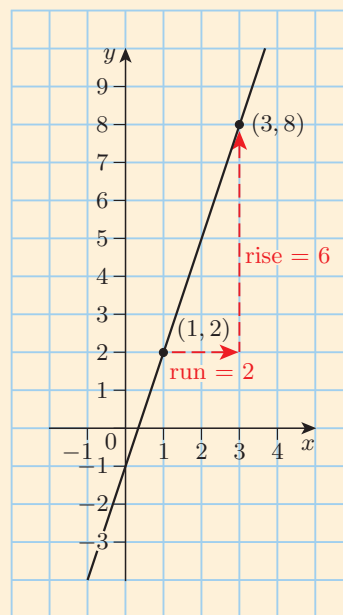


Figure 6 The run and rise between the points $(1, 2)$ and $(3, 8)$

To see why this method works, notice that saying that your pencil tip moves six units up for every two units right is the same as saying that it moves three units up for every one unit right.

The method is summarised below.

Strategy To calculate the gradient (slope) of a straight line

Choose two points on the line. Then

$$\text{gradient} = \frac{\text{increase in } y\text{-coordinate}}{\text{increase in } x\text{-coordinate}} = \frac{\text{rise}}{\text{run}}.$$

(The run and rise are calculated from the *scales on the axes*, not from the physical distances on the paper or screen.)

Try to remember that the gradient is the increase in the y -coordinate divided by the increase in the x -coordinate, not the other way round. You might like to remember this by the fact that you can pronounce yox (y over x) but not xoy (x over y)!

Some lines, like the one in Figure 7, slope down rather than up from left to right. As you move your pencil tip from a left-hand point to a right-hand point along a line like this, the y -coordinate decreases rather than increases. A decrease can be thought of as a ‘negative increase’, so this means that the rise is negative.

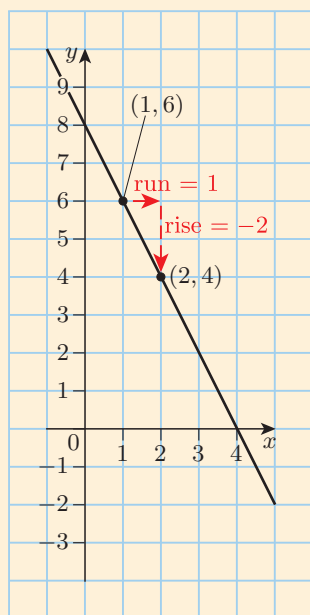


Figure 7 A negative rise

For example, consider the two points plotted in Figure 7. As your pencil tip moves from the left-hand point $(1, 6)$ to the right-hand point $(2, 4)$, the x -coordinate increases from 1 to 2, which is an increase of 1, and the y -coordinate decreases from 6 to 4, which is a decrease of 2, or an increase of -2 . So the run is 1 and the rise is -2 . The gradient is therefore

$$\frac{\text{rise}}{\text{run}} = \frac{-2}{1} = -2.$$

Any line that slopes down from left to right has a negative gradient, since the y -coordinate decreases as the x -coordinate increases.

The sign of the gradient

Lines that slope up from left to right have a *positive* gradient.

Lines that slope down from left to right have a *negative* gradient.

When you calculate the gradient of a line, you should always check that the sign of the gradient agrees with the direction in which the line slopes.

There is a systematic way to work out the run and rise between two points on a line. Since the run is the increase in the x -coordinate, you can calculate it by subtracting the x -coordinate of the left-hand point from the x -coordinate of the right-hand point. Similarly, since the rise is the increase in the y -coordinate, you can calculate it by subtracting the y -coordinate of the left-hand point from the y -coordinate of the right-hand point, and this works even if the rise is negative. For example, for the points $(1, 6)$ and $(2, 4)$ in Figure 7, the run is

$$2 - 1 = 1$$

and the rise is

$$4 - 6 = -2,$$

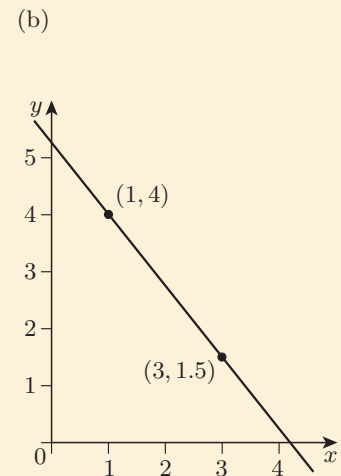
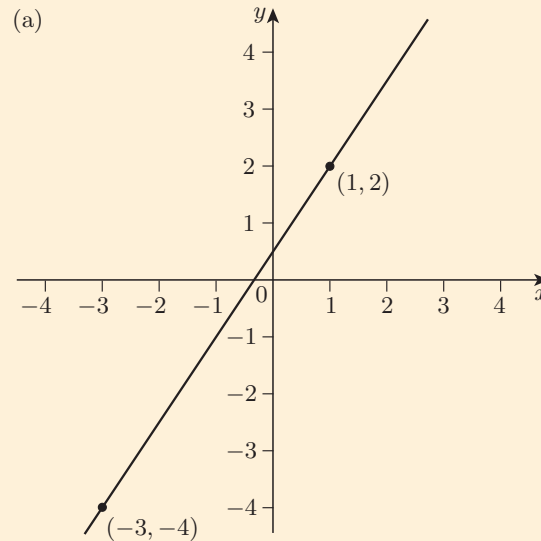
as found above. This way of working out the run and rise is used in the next example.



Tutorial clip

Example 3 Calculating gradients of lines

Calculate the gradients of the lines shown below.

**Solution**

- (a) The run is the increase in the x -coordinate, which is $1 - (-3) = 4$.

The rise is the increase in the y -coordinate, which is $2 - (-4) = 6$.

So the gradient is $\frac{\text{rise}}{\text{run}} = \frac{6}{4} = 1.5$.

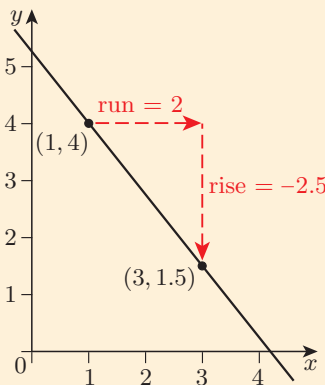
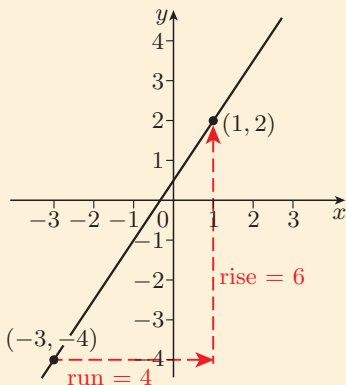
(Check: The line slopes up, so the gradient should be positive, which it is.)

- (b) The run is the increase in the x -coordinate, which is $3 - 1 = 2$.

The rise is the increase in the y -coordinate, which is $1.5 - 4 = -2.5$.

So the gradient is $\frac{\text{rise}}{\text{run}} = \frac{-2.5}{2} = -1.25$.

(Check: The line slopes down, so the gradient should be negative, which it is.)



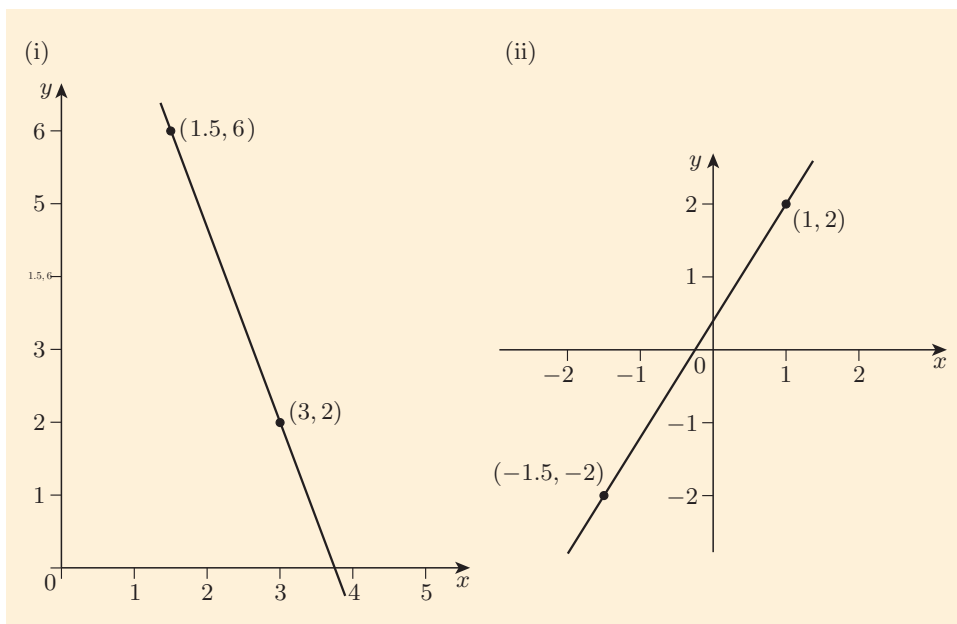
Recall that a *terminating* decimal is one that has a finite number of digits after the decimal point.

The gradients calculated in Example 3 are expressed as decimals, but you can also express gradients as fractions. If the decimal form of a gradient does not terminate, then it is usually best to express it as a fraction, if possible, so that it is an exact value. For example, if the rise is 2 and the run is 3, then the gradient should be written as $\frac{2}{3}$, rather than a rounded decimal such as 0.67.

The next activity asks you to calculate the gradients of some lines. Although you can use any two points on a line to find the gradient, if a question gives you the coordinates of two points on a line, then you should use these points rather than reading off new coordinates from a graph, as your readings may not be accurate.

Activity 6 Calculating gradients of lines

(a) Calculate the gradients of the lines shown below.



(b) For each of the following pairs of points, calculate the gradient of the line that passes through them.

(i) $A(2, 2)$ and $B(-2, 2)$

(ii) $A(2, 2)$ and $C(-1.5, -2.5)$

(iii) $A(2, 2)$ and $D(3, -1)$

(You might find it helpful to first sketch these points on a graph.)

(c) What is the rise between any two points on a *horizontal* line? What is the gradient of a horizontal line?

When a line is drawn on a graph, how steep it looks on the page or screen depends not only on its gradient but also on the scales used on the axes. For example, the three graphs in Figure 8 all show the line passing through the points $(1, 2)$ and $(5, 7)$, so the gradient is the same in all three cases (namely $\frac{5}{4}$), but the steepnesses of the lines look different.

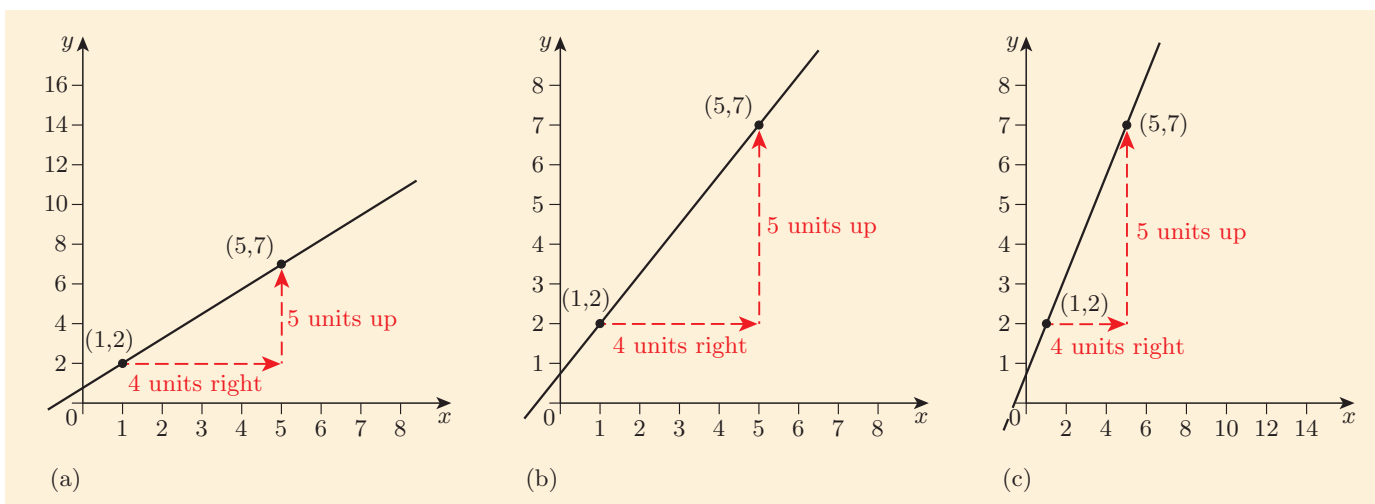


Figure 8 The effect of changing the axis scales

Because of this, when you draw a graph using the standard x - and y -axes, it is usually a good idea to use the same scale on each axis. Figure 9 shows lines with gradients 1 and -1 drawn on graphs with the same scale on each axis. As you can see, these lines make angles of 45° with the x -axis.

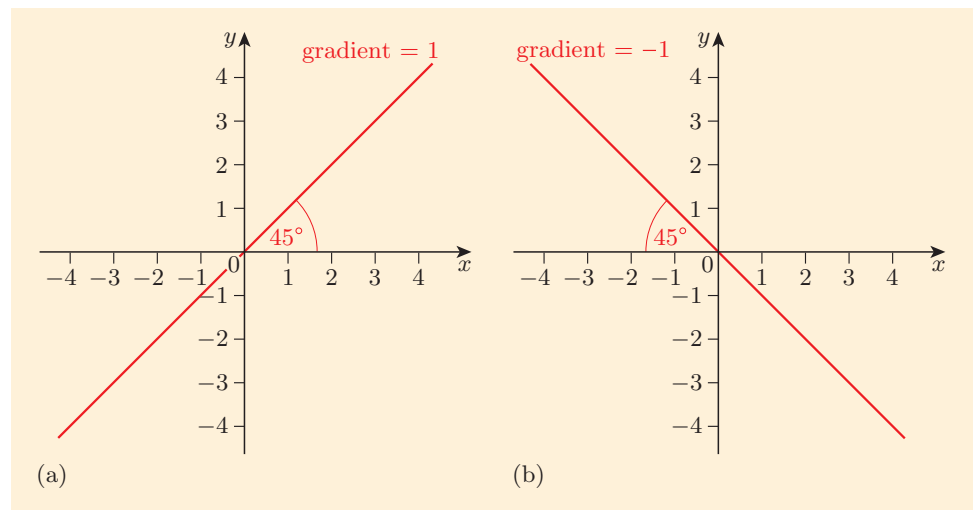


Figure 9 Lines with gradients 1 and -1 drawn on axes with equal scales

The gradients of other lines drawn on graphs with the same scale on each axis can be compared to these gradients, as follows.

- A line whose gradient is greater than 1 makes an angle greater than 45° with the x -axis. So does a line whose gradient is greater than 1 in size, but negative. For example, the blue lines in Figure 10 have gradients 3 and -3 and make angles greater than 45° with the x -axis.
- A line whose gradient is smaller than 1, but positive, makes an angle less than 45° with the x -axis. So does a line whose gradient is smaller than 1 in size, but negative. For example, the green lines in Figure 10 have gradients $\frac{1}{3}$ and $-\frac{1}{3}$ and make angles less than 45° with the x -axis.

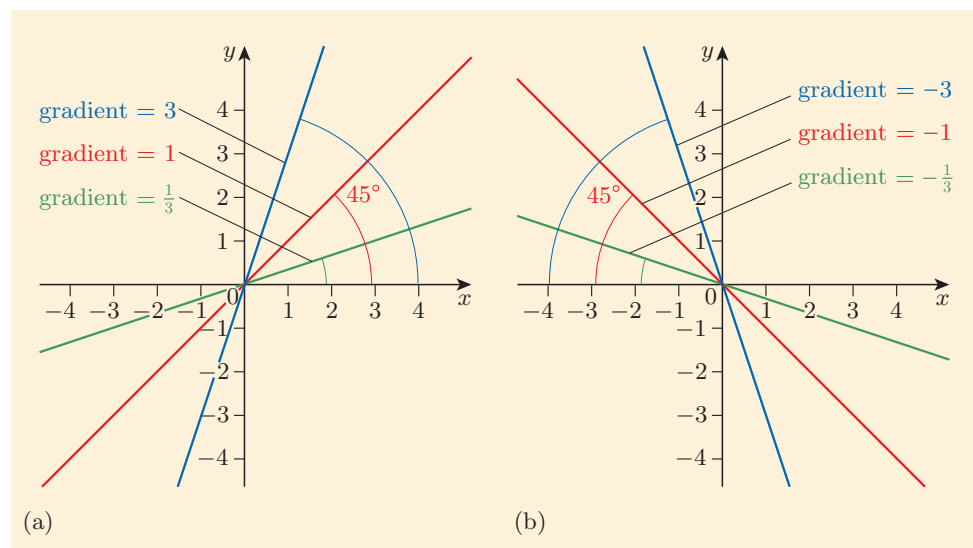


Figure 10 (a) Positive gradients of lines and angles made with the x -axis. (b) Negative gradients of lines and angles made with the x -axis.

When you calculate the gradient of a line that has been drawn using axes with equal scales, it is a good idea to make sure that the size of the gradient seems to agree with the angle that the line makes with the x -axis.

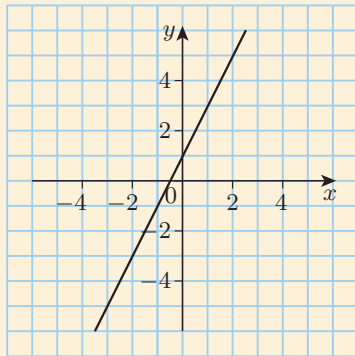
By the **size** of a number we mean its value without its negative sign, if it has one. For example, the size of 3 is 3 and the size of -3 is also 3. The size of a number is often referred to as its *modulus*, *magnitude* or *absolute value*.

Activity 7 Estimating the gradients of graphs

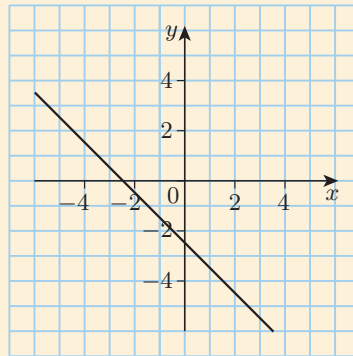
For each of the graphs below, decide how the angle that the line makes with the x -axis compares with 45° , and hence state which of the following apply to the gradient of the line:

about 1, about -1 , greater than 1, less than -1 ,
between 0 and 1, between -1 and 0.

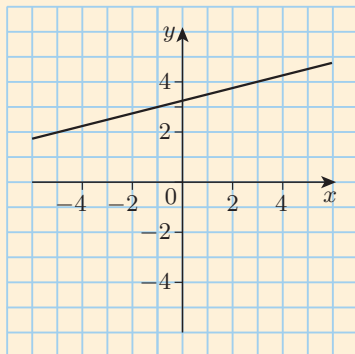
(a)



(b)



(c)



(d)

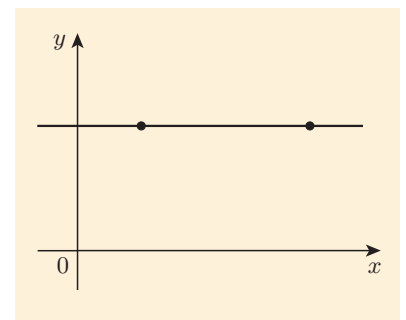
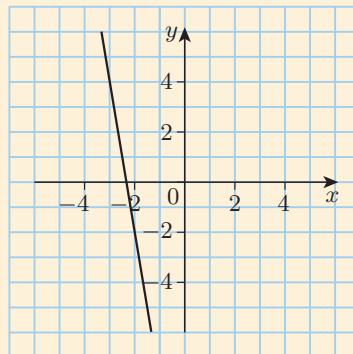


Figure 11 The rise between two points on a horizontal line is zero

Horizontal and vertical lines

You saw in Activity 6(c) that the gradient of every horizontal line is zero. This is because the gradient of a line is the rise divided by the run, and the rise between any two points on a horizontal line is zero, as shown in Figure 11.

What about vertical lines? For example, consider the line passing through the points $(1, 1)$ and $(1, 4)$, which is shown in Figure 12.

Here we do not have a left-hand point and a right-hand point, and whichever way round we take the points, the run is zero. Since the gradient of a line is the rise divided by the run, and it is not possible to divide a number by zero, it is not possible to calculate the gradient of this line. The same is true for every vertical line. So the gradient of a vertical line is undefined.

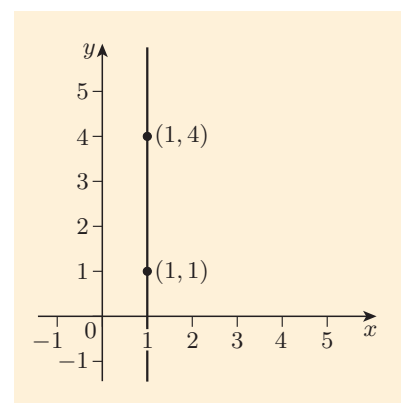


Figure 12 A vertical line

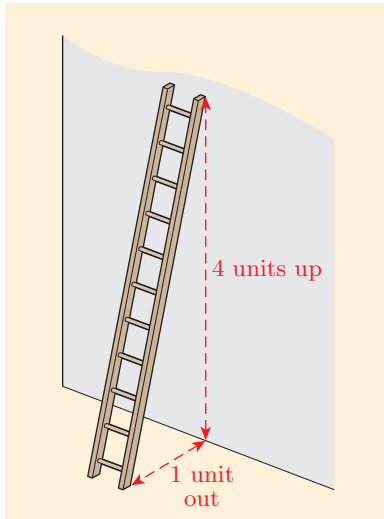


Figure 13 The recommended slope for a ladder is 1 unit out and 4 units up, which is a gradient of $\frac{4}{1} = 4$

x_1 is read as 'x-one'.

There are many situations where it is important to know how steep something is. For example, you may need to decide whether a ladder has been put up safely (Figure 13), determine how steep a hill is on a proposed walk, or even decide whether an avalanche is likely given the slope of a snowfield.

2.2 A formula for gradient

In the previous subsection you saw a method for calculating the gradient of a line, using the coordinates of two points on the line. If letters are used to represent the coordinates of the points, then this method can be summarised as a formula.

Since coordinates are written in the form (x, y) , and there are two points, it is convenient to represent the coordinates of the left-hand point by (x_1, y_1) and those of the right-hand point by (x_2, y_2) , as shown in Figure 14. The small numbers after the x and y are known as **subscripts**. They are labels that distinguish one point from another.

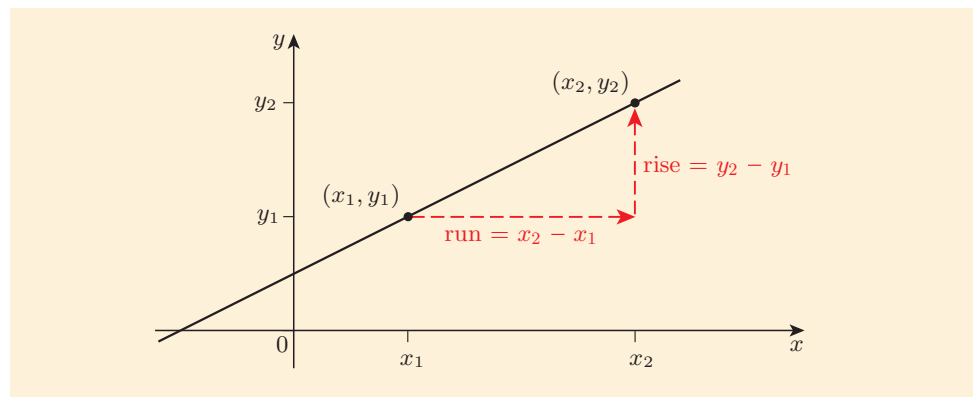


Figure 14 The run and rise between the points (x_1, y_1) and (x_2, y_2)

The run is calculated by subtracting the x -coordinate of the left-hand point from the x -coordinate of the right-hand point, so

$$\text{run} = x_2 - x_1.$$

Similarly, the rise is calculated by subtracting the y -coordinate of the left-hand point from the y -coordinate of the right-hand point, so

$$\text{rise} = y_2 - y_1.$$

Hence

$$\text{gradient} = \frac{\text{rise}}{\text{run}} = \frac{y_2 - y_1}{x_2 - x_1}.$$

This formula is used to calculate the gradient of a line in the next example.



Tutorial clip

Example 4 Using the formula for gradient

Use the formula

$$\text{gradient} = \frac{y_2 - y_1}{x_2 - x_1},$$

to find the gradient of the line through the points $(1, 1)$ and $(3, -1)$.

Solution

A sketch of the line is shown in the margin.

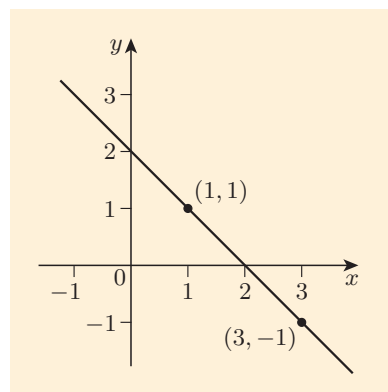
The left-hand point is $(1, 1)$, so $x_1 = 1$ and $y_1 = 1$.

The right-hand point is $(3, -1)$, so $x_2 = 3$ and $y_2 = -1$.

The formula gives

$$\text{gradient} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{-1 - 1}{3 - 1} = \frac{-2}{2} = -1.$$

(Check: The line slopes down, so the gradient should be negative, which it is. Also, the same scale has been used on each axis in the diagram, and the line seems to make an angle of about 45° with the x -axis, so the size of the gradient should be about 1, which it is.)



In Example 4, the formula

$$\text{gradient} = \frac{y_2 - y_1}{x_2 - x_1} \quad (1)$$

was used to work out the gradient of a line, with $(x_1, y_1) = (1, 1)$ and $(x_2, y_2) = (3, -1)$. What happens if you take the points the other way round? With $(x_1, y_1) = (3, -1)$ and $(x_2, y_2) = (1, 1)$, the formula gives

$$\text{gradient} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{1 - (-1)}{1 - 3} = \frac{2}{-2} = -1.$$

This is the same value as found in Example 4, so here it does not matter whether (x_1, y_1) is taken to be the left-hand point and (x_2, y_2) the right-hand point, or the other way round. But is this true in general?

You know that formula (1) holds if (x_1, y_1) is the left-hand point and (x_2, y_2) the right-hand point, because that was worked out at the beginning of this subsection. To check whether it holds if (x_1, y_1) is the right-hand point and (x_2, y_2) the left-hand point, let's use the usual method to work out the gradient in this case. The run is calculated by subtracting the x -coordinate of the left-hand point from the x -coordinate of the right-hand point, so

$$\text{run} = x_1 - x_2.$$

Similarly, the rise is calculated by subtracting the y -coordinate of the left-hand point from the y -coordinate of the right-hand point, so

$$\text{rise} = y_1 - y_2.$$

Hence

$$\text{gradient} = \frac{\text{rise}}{\text{run}} = \frac{y_1 - y_2}{x_1 - x_2}. \quad (2)$$

This formula looks different from formula (1), but it is just the same formula in disguise. If you multiply top and bottom of the fraction in formula (2) by -1 , then you obtain

$$\begin{aligned} \text{gradient} &= \frac{(-1) \times (y_1 - y_2)}{(-1) \times (x_1 - x_2)} \\ &= \frac{-y_1 + y_2}{-x_1 + x_2} \\ &= \frac{y_2 - y_1}{x_2 - x_1}, \end{aligned}$$

which is the same as formula (1).

Remember that multiplying both top and bottom of a fraction by a non-zero number does not change the value of the fraction.

So the result that you have seen in this subsection can be summarised as follows.

A formula for gradient

The gradient of the line through the points (x_1, y_1) and (x_2, y_2) is given by

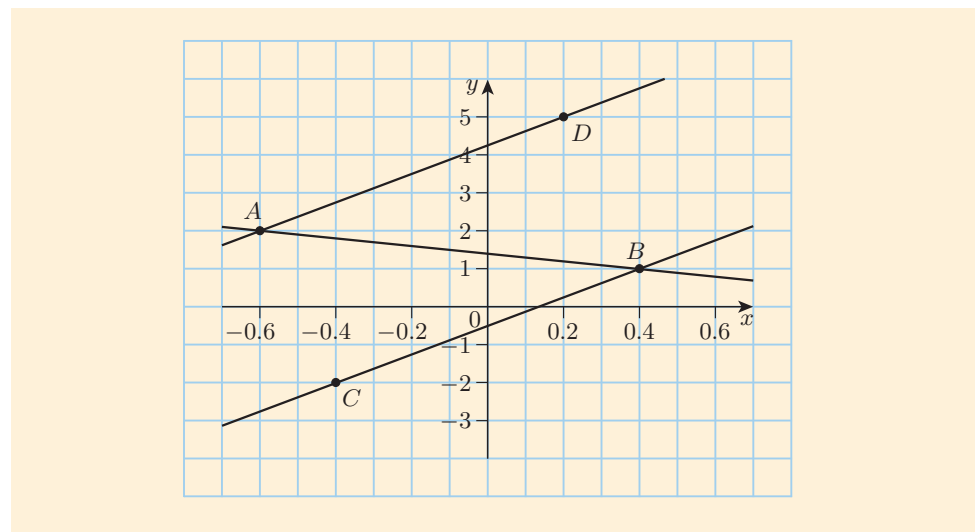
$$\text{gradient} = \frac{y_2 - y_1}{x_2 - x_1}.$$

(It does not matter which point you take to be (x_1, y_1) and which you take to be (x_2, y_2) .)

In the next activity you are asked to use this formula to calculate some gradients. Notice that the graph in this activity has been drawn with different scales on the x - and y -axes. This is because it would be difficult to read off the coordinates of the points if the graph were drawn with equal scales.

Activity 8 Using the formula for gradient

Consider the following graph.



- Write down the coordinates of the points A , B , C and D .
- Use the formula for gradient to calculate the gradients of the lines that pass through the following pairs of points.
 - A and B
 - A and D
 - B and C
- What do you notice about the gradients of the lines in parts (b)(ii) and (b)(iii)?

Two lines are said to be **parallel** if they never cross, even when extended infinitely far in each direction. For example, in the graph in Activity 8, the line through the points A and D is parallel to the line through the points B and C . You saw in the solution to part (c) of the activity that these two lines have the same gradient. In general, saying that two lines are parallel means the same as saying that they have the same gradient (except for vertical lines, which are parallel but whose gradient is undefined).

2.3 Interpreting gradient

In this subsection you will see some real-life examples of graphs and learn how gradients can be interpreted practically.

A straight-line graph that illustrates a relationship between real-life quantities usually has axis scales that represent particular units. So the gradient of the line also has units, which should be quoted when the gradient is stated. The units of the gradient are the units on the vertical axis divided by the units on the horizontal axis, since this corresponds to the formula for gradient.

For example, the graph in Figure 15 illustrates the growth of a bamboo plant.

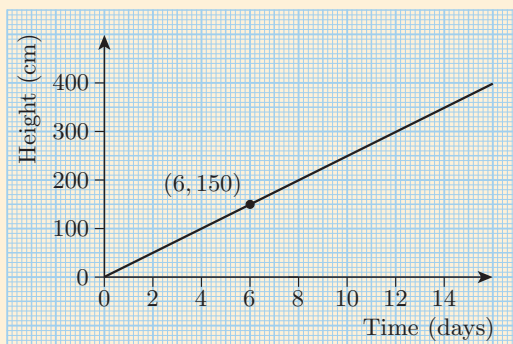


Figure 15 The growth of a bamboo plant

The vertical axis measures the height of the bamboo plant in centimetres, and the horizontal axis measures the time in days. So the units for the gradient are cm/day, or centimetres per day.

The line on the graph passes through the points (0,0) and (6,150), so the numerical value of the gradient is

$$\frac{150 - 0}{6 - 0} = \frac{150}{6} = 25,$$

and hence the gradient is 25 cm/day. This quantity measures the number of centimetres that the bamboo plant grows in 1 day.

In other words, the gradient gives the *rate* at which the bamboo grows. In general, the gradient of a graph tells you the amount that the variable on the vertical axis changes when the variable on the horizontal axis increases by one unit. So if the variables on the horizontal and vertical axes are x and y respectively then the gradient is the **rate of change** of y with respect to x .

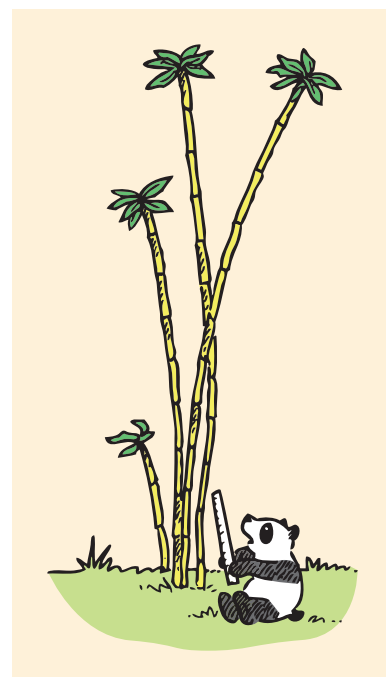
To save time when doing calculations like the one above, you can do the calculation and quote the units at the same time, in the following way.

The gradient is

$$\frac{150 - 0}{6 - 0} = \frac{150}{6} = 25 \text{ cm/day}.$$

It is not strictly correct to write this, since whatever is on the left of an equals sign should be exactly equal to whatever is on the right – so if units are included on one side, then they should also be included on the other side. However, calculations can look unnecessarily complicated if units are included all the way through, so in practice it is sometimes convenient to omit them until the final answer is obtained.

If the units on the two axes are *the same*, then when you divide one by the other they cancel out and so the gradient has no units.



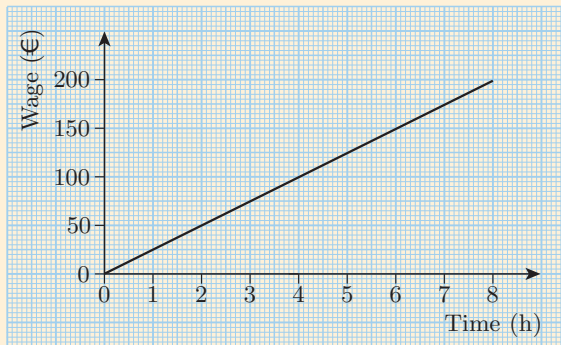
Some types of bamboo grow at about a metre per day in favourable conditions.

Here are some more examples of interpreting gradients in practical situations.

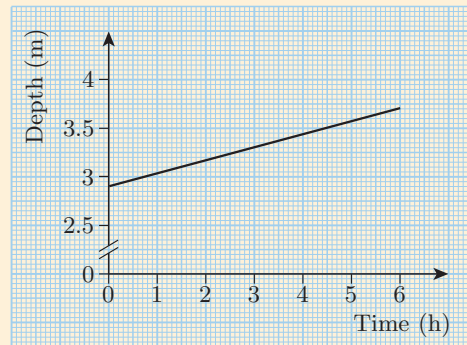
Example 5 Interpreting gradients

Graph (a) below shows the wages earned by a factory worker for shifts of different lengths, and graph (b) shows the depth of a river over a 9-hour period after heavy rainfall. For each graph, find the gradient and explain what it measures.

(a)



(b)



Solution

- (a) Choose two points on the line whose coordinates can be easily read off the graph.

The points (0, 0) and (4, 100) lie on the line, so its gradient is

$$\frac{100 - 0}{4 - 0} = \frac{100}{4} = 25 \text{ euro/hour.}$$

The gradient measures the number of € earned by the worker per hour.

- (b) The points (0, 2.9) and (6, 3.7) lie on the line, so its gradient is

$$\frac{3.7 - 2.9}{6 - 0} = \frac{0.8}{6} = 0.13 \text{ m/h (to 2 d.p.).}$$

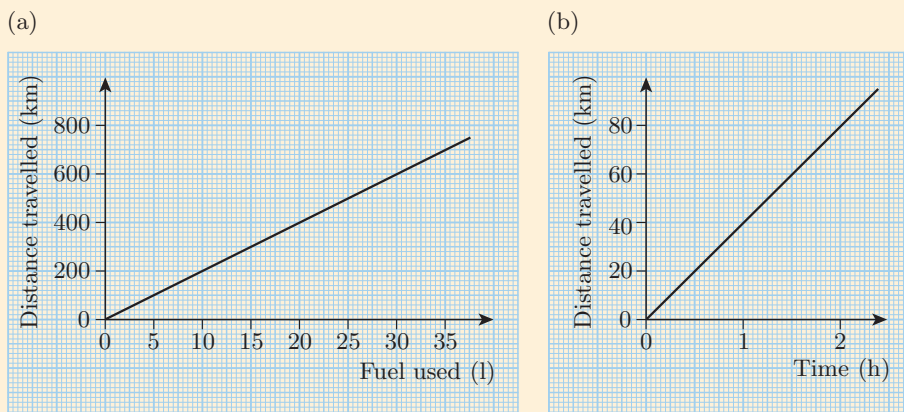
The gradient measures the number of metres by which the river rises per hour.



Figure 16 A gauge used to measure the depth of a river

Activity 9 Interpreting gradients

Graph (a) on the next page shows the distance travelled by a car plotted against the amount of fuel used, and graph (b) shows the distance travelled plotted against the time taken. For each graph, find the gradient and explain what it measures.



Remember that 'l' stands for litres.

Figure 17 is an example of a real-life graph with a negative gradient. It shows the depth of the river in Example 5(b) over an earlier 10-hour period.

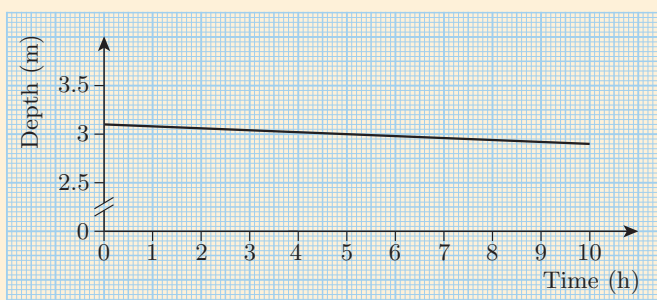


Figure 17 The depth of a river over a 10-hour period

The points (0, 3.1) and (10, 2.9) lie on the line, so its gradient is

$$\frac{2.9 - 3.1}{10 - 0} = \frac{-0.2}{10} = -0.02 \text{ m/h.}$$

The fact that the gradient is negative tells you that the depth of the river *decreases* with time during this period. Its depth decreases by 0.02 metres per hour, or 2 centimetres per hour. The box below summarises what the sign of the gradient of a graph tells you.

Interpreting the sign of the gradient of a graph

- A *positive* gradient indicates that the quantity on the vertical axis *increases* as the quantity on the horizontal axis increases.
- A *negative* gradient indicates that the quantity on the vertical axis *decreases* as the quantity on the horizontal axis increases.
- A *zero* gradient indicates that the quantity on the vertical axis *remains constant* as the quantity on the horizontal axis increases.

Sometimes when you want to model a situation, it is helpful to use a graph that consists of more than one straight line. For example, consider the graph in Figure 18. It shows the depth of the river in Example 5(b), over



an extended period that includes the 6-hour period in the example, the 10-hour period in Figure 17, and another 4-hour period in between.

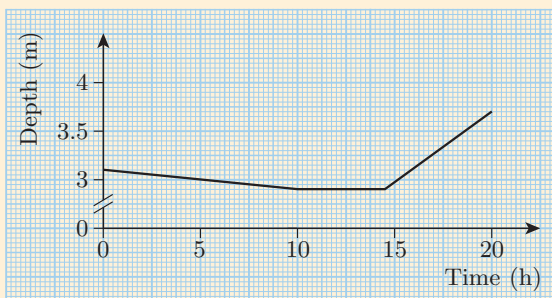


Figure 18 The depth of a river over a 20-hour period

The graph shows that in the first 10 hours the depth of the river fell from 3.1 m to 2.9 m, over the next four hours it remained constant at 2.9 m, and over the final six hours, after the heavy rain, its depth increased to 3.7 m. The gradients of the three line segments that make up the graph can be calculated in the usual way. Each of these gradients indicates the rate of change of the depth of the river with respect to time during one of the three time periods.

It is likely that the graph in Figure 18 is a simplified model, perhaps based on only four measurements, at 0 hours, 10 hours, 14 hours and 20 hours. If more measurements had been taken, then it might have been possible to use a curved graph to provide a more accurate model. However the graph in Figure 18 is sufficient for many purposes.

2.4 Intercepts

The value on the axis scale where a straight-line graph crosses one of the graph axes is another important characteristic of the graph, and is called an **intercept**. In particular, if the graph is drawn using the standard x - and y -axes, then the x -intercept is the value where the line crosses the x -axis, and the y -intercept is the value where the line crosses the y -axis. In other words, the x -intercept is the value of x when $y = 0$, and the y -intercept is the value of y when $x = 0$.

For example, the x - and y -intercepts of the line shown in Figure 19 are -3 and 2 , respectively. Notice that an intercept is a value and not a point. The points corresponding to these intercepts are $(-3, 0)$ and $(0, 2)$.

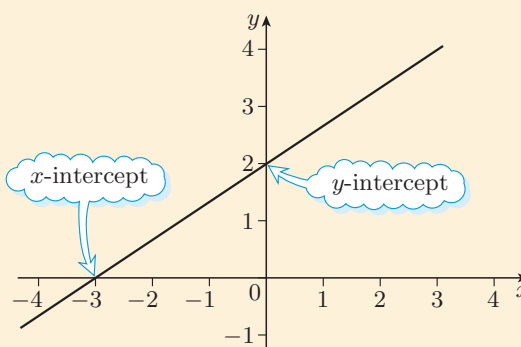


Figure 19 The x - and y -intercepts of a line

Note that the word ‘intercept’ is different from ‘intersect’. Two lines *intersect* if they cross each other.

If the part of a line shown on a graph does not cross an axis, then it may be necessary to extend the line to find the intercept.

As with some other terms used for graphs, the terms used for intercepts are adjusted according to the axis labels. For example, if a graph has a c -axis and an f -axis, then it has a c -intercept and an f -intercept. You can also refer to horizontal and vertical intercepts.

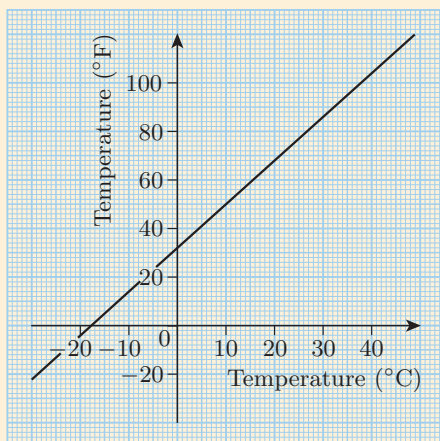
If a line is drawn on a graph whose axis scales represent particular units, then the intercepts also have units, which should be quoted when the intercepts are stated.

The next example illustrates how some intercepts can be interpreted practically.

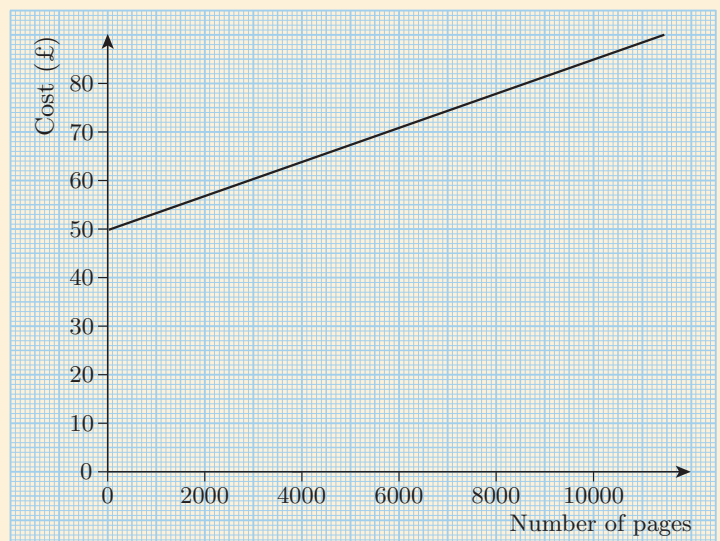
Example 6 Interpreting vertical intercepts

Graph (a) below is the Celsius–Fahrenheit conversion graph, and graph (b) shows the cost of printing pages on a home printer. For graph (a), write down the values of the intercepts, and state what they represent. Repeat this for graph (b), but only for the vertical intercept.

(a)



(b)



Solution

- (a) The vertical intercept is 32°F . This is the temperature in $^{\circ}\text{F}$ when the temperature in $^{\circ}\text{C}$ is zero.

The horizontal intercept is about -18°C . This is the temperature in $^{\circ}\text{C}$ when the temperature in $^{\circ}\text{F}$ is zero.

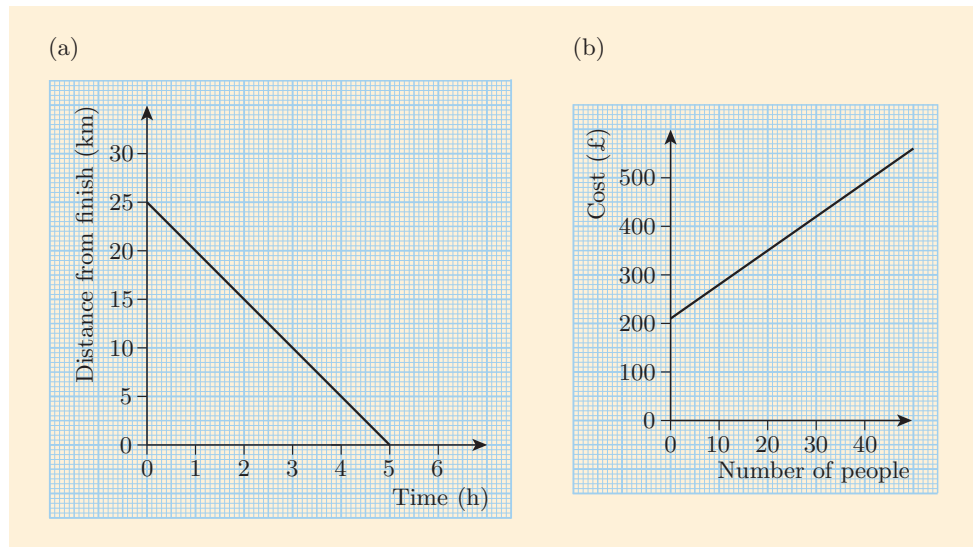
- (b) The vertical intercept is $\pounds 50$. This is the cost when the number of pages printed is zero. That is, it is the initial cost of the printer.

You were asked to find this value earlier, in Activity 4(d)(iv).

The next activity asks you to find some intercepts.

Activity 10 *Interpreting intercepts*

Graph (a) below shows the distance of a participant in a sponsored walk from the finish line at different times after the start time, and graph (b) shows the cost of hiring a particular venue. For graph (a), write down the values of the intercepts, and state what they represent. Repeat this for graph (b), but only for the vertical intercept.



In this section, you have learned about the gradient and the intercepts of a straight line. The gradient measures the rate at which the vertical variable changes with respect to the horizontal variable, and you have learned how to calculate it from two points on the line. The intercepts are the scale values where the line crosses the axes. You have seen how to interpret both the gradient and the intercepts in practical situations.

3 Equations of straight lines

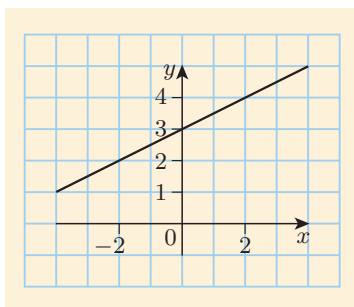


Figure 20 The line
 $y = \frac{1}{2}x + 3$

In Example 1 on page 65 you saw that the line shown in Figure 20 is described by the equation $y = \frac{1}{2}x + 3$. The points that lie on the line are the points whose coordinates (x, y) satisfy this equation.

In this section you will see that every straight line that can be drawn on a graph is described by an equation in a similar way. As usual, the ideas are explained using the standard variables x and y , but they apply to any pair of variables whose relationship is illustrated by a straight line on a graph.

The first subsection is about straight lines of a particular type – those that pass through the origin.

3.1 Lines that pass through the origin

Consider the line shown in Figure 21, which passes through the origin and has gradient 2.

Since the gradient is 2, if you position your pencil tip on the line and then move it a number of units right and twice that number of units up, then your pencil tip ends up at a point that is also on the line. So, since the origin is on the line, so are the points $(1, 2)$, $(3, 6)$ and $(4.5, 9)$, as shown in Figure 21, and so is any point with positive coordinates whose y -coordinate is twice its x -coordinate.

Similarly, if you position your pencil tip at a point on the line and move it a number of units to the left and twice that number of units down, then it ends up at another point on the line. So the points $(-1, -2)$ and $(-3, -6)$ also lie on the line, and so does any point with negative coordinates whose y -coordinate is twice its x -coordinate.

So every point whose y -coordinate is twice its x -coordinate lies on the line.

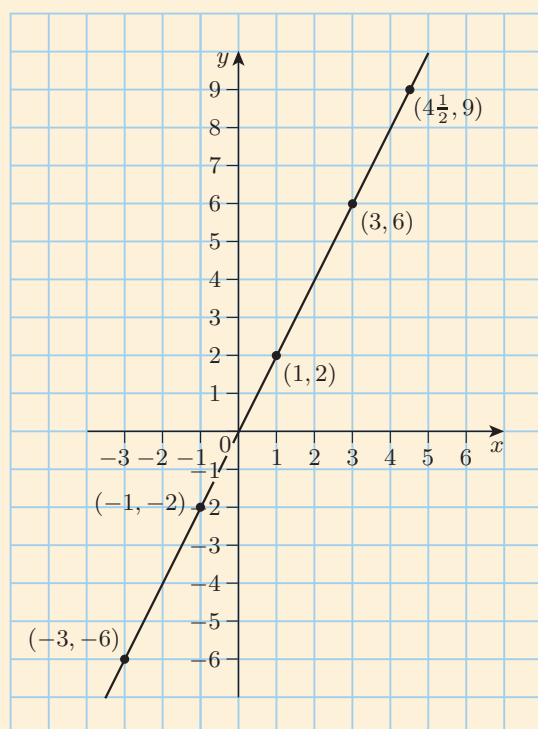
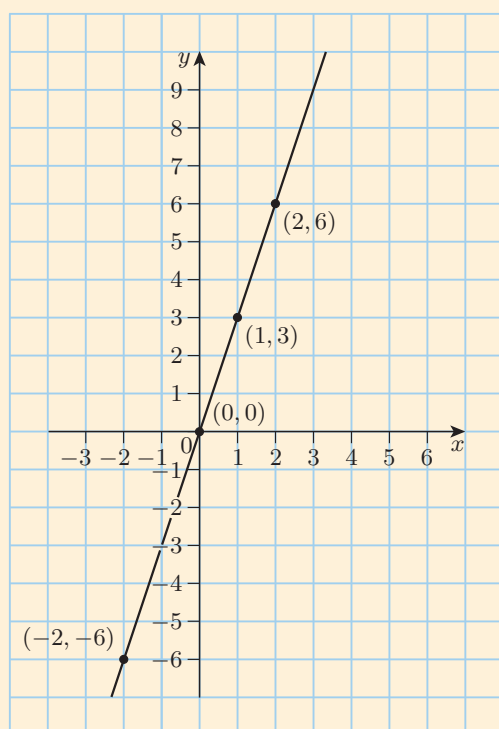


Figure 21 The line through the origin with gradient 2

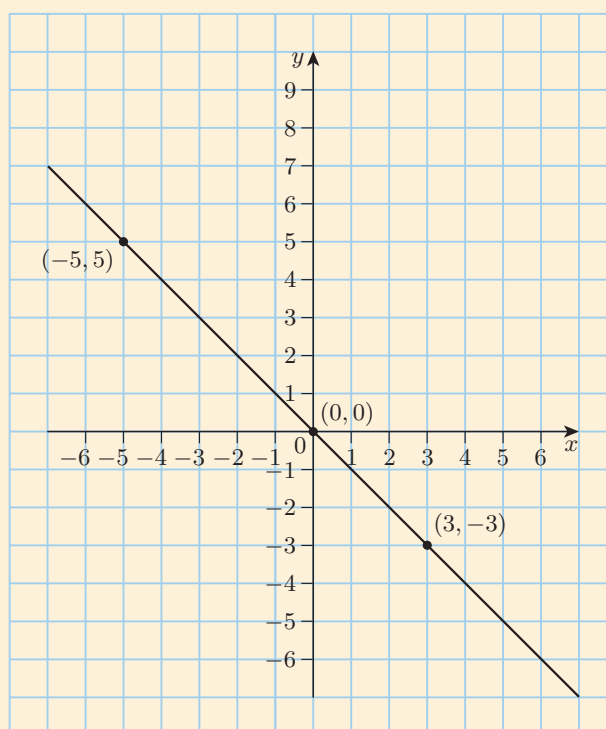
You can also see that every point whose y -coordinate is *not* twice its x -coordinate does *not* lie on the line. For example, consider the point $(5, 3)$: the number 3 is not 2 times 5, and this point does not lie on the line.

Since the points that lie on the line are the points such that the y -coordinate is twice the x -coordinate, the equation of the line is $y = 2x$.

Other lines that pass through the origin have similar equations. For example, Figure 22(a) shows the line with gradient 3 that passes through the origin. The y -coordinate of every point on this line is three times its x -coordinate, so the equation of this line is $y = 3x$. Similarly, Figure 22(b) shows the line with gradient -1 that passes through the origin. The y -coordinate of every point on the line is -1 times its x -coordinate, so the equation of this line is $y = -1x$, or $y = -x$.



(a)



(b)

Figure 22 (a) The line through the origin with gradient 3. (b) The line through the origin with gradient -1 .

In general, each line that passes through the origin has the equation

$$y = \text{gradient} \times x,$$

The equations of vertical lines are discussed later in the section.

with the exception of the vertical line through the origin (the y -axis) because it does not have a gradient. This fact can be stated as follows.

The equation of a line through the origin

The straight line that passes through the origin and has gradient m has equation $y = mx$.

It is traditional to use the letter m to represent gradient, though the reason is no longer known! The earliest known use of the letter m for gradient is by the Italian mathematician Vincenzo Riccati in 1757.

In the next activity you are asked to use Graphplotter, a computer tool that draws graphs, to view some more lines of the form $y = mx$, and to convince yourself that the gradient of each line is m , the coefficient of x . There are instructions for using Graphplotter in the MU123 Guide.

You will notice that Graphplotter uses the word ‘function’. Whenever you have an equation that expresses one variable in terms of another variable, you can think of it as a rule that takes an input value and produces an output value. For example, if the equation is $y = 2x$, then inputting $x = 2$ gives the output $y = 4$, inputting $x = 3$ gives the output $y = 6$, and similarly for other values. A rule that takes input values and produces output values like this is often called a **function**.

So the graph of an equation of the form $y = mx$ is the graph of a function.

Activity 11 Investigating lines of the form $y = mx$ 

Graphplotter

- Open Graphplotter and choose the equation $y = mx + c$ from the drop-down list. Set $c = 0$, since this activity is about equations of the form $y = mx$.
- Use the slider to increase the value of m gradually up to 10, and then down again, to -10 . Observe how the gradient of the line changes as you change the value of m .
- Use Graphplotter to plot the lines with equations $y = 3x$, $y = 0.5x$, $y = -2x$ and $y = -5x$, in turn, by typing the appropriate values of m into the box and pressing 'Enter'. (Do not change the axis scales – keep them the same for each graph.) Which of these lines makes the smallest angle with the x -axis?

Later in this section you will see how other types of lines – those that do not pass through the origin – are described by equations. First, however, in the next subsection you will learn about a commonly-occurring type of relationship between two quantities, and you will see that relationships of this type are illustrated by straight lines through the origin.

3.2 Direct proportion

Sometimes two quantities are related in such a way that if you multiply or divide one of the quantities by a number, then the other quantity is also multiplied or divided by the same number. For example, the relationship between distance in miles and the equivalent distance in kilometres is like this. If you double the number of miles, then the number of kilometres is also doubled. And if you halve the number of miles, then the number of kilometres is also halved, and so on.

If two quantities are related in this way, then they are said to be *directly proportional* to each other (or just proportional to each other, for short), and the relationship is known as **direct proportion**. So, for example, the number of miles is directly proportional to the equivalent number of kilometres.

Many everyday relationships between two quantities are direct proportion relationships. Here are two more examples.

- The number of people attending a concert is directly proportional to the money received from the ticket sales (if all the tickets have the same price).
- The volume of water that you need to boil to make tea is directly proportional to the number of cups of tea that you want to make (provided that each cup is to be filled with the same amount of tea).

On the other hand, there are many relationships between two quantities that are *not* direct proportion relationships. Here is an example.

- The speed at which you travel during a particular journey is *not* directly proportional to the time that the journey takes. For example, suppose that you have to make a 40-kilometre journey. If you travel at 40 km/h, then the time taken is 1 hour; but if you travel at 80 km/h, then the time taken is not 2 hours!

Sometimes a relationship that you might expect to be direct proportion is in fact not. For example, if you were buying some items, then you might expect to pay twice the price if you buy twice as many items. But in practice, twice the number of items may cost less than twice the price – there may be a discount, or a special offer such as ‘buy one, get one free’.

Activity 12 Identifying direct proportion relationships

In each of parts (a) to (e), state whether the two quantities are directly proportional.

In each case, ask yourself: If I double the first quantity (or multiply or divide it by any number), is the second quantity also doubled (or multiplied or divided by the same number)?

- The time for which you travel and the distance that you travel, if you are travelling at a constant speed.
- The number of painters and the time it takes them to paint a bridge, if they all help and all work at the same rate.
- The number of pounds that you exchange for euros and the number of euros that you receive, if there is a transaction fee of €10.
- The number of songs that you download from a music website and the total cost, if each song costs the same amount.
- The number of hours that you work and the gross pay that you receive, if you are paid at a fixed hourly rate.

Gross pay is pay before deductions such as taxes have been made.

You saw earlier in this subsection that the relationship between distance in miles and the equivalent distance in kilometres is a direct proportion relationship. This relationship is described by the formula

$$K = 1.6M,$$

where K and M represent the number of kilometres and the number of miles, respectively. (This formula is only approximate.)

Whenever two quantities are directly proportional, their relationship is described by a formula similar to this one.

Direct proportion relationships

If two quantities x and y are directly proportional to each other, then the relationship between them is described by an equation of the form

$$y = kx,$$

where k is a non-zero number, known as the **constant of proportionality**.

The statement ‘ y is directly proportional to x ’ is sometimes written as

$$y \propto x.$$

A more precise, but still not exact, formula is $K = 1.609M$.

The symbol \propto is read as ‘is proportional to’.

A **constant** in an equation or expression is a quantity that does not change when the values of the variables change. Sometimes, as here, it is convenient to represent a constant by a letter.

If two variables are directly proportional and you know a value of one variable and the corresponding value of the other variable, then you can

work out the constant of proportionality. The next example shows you how to do this.

Example 7 Finding a constant of proportionality

The two quantities x and y are directly proportional to each other, and $y = 24$ when $x = 16$.

- (a) Find a formula for y in terms of x .
- (b) Find the value of y when $x = 20$.

Solution

- (a) Since x and y are directly proportional, their relationship is expressed by an equation of the form

$$y = kx,$$

where k is a constant.

 Use the given values of x and y to find the value of k . 

Also, $y = 24$ when $x = 16$. Substituting these values into the equation gives

$$24 = 16k.$$

We now solve this equation to find the value of k .

$$\text{Divide by 16:} \quad \frac{24}{16} = k$$

$$\text{Simplify the fraction and swap the sides:} \quad k = \frac{3}{2}$$

Substituting $k = \frac{3}{2}$ into the equation $y = kx$ gives

$$y = \frac{3}{2}x.$$

This is the formula required.

- (b) Substituting $x = 20$ into the formula found in part (a) gives

$$y = \frac{3}{2} \times 20 = 30.$$

Finding a constant of proportionality can be useful in many different practical situations – here is an example.

An art installation called ‘Of all the people in all the world’ has been exhibited at various venues throughout the world in recent years. It aims to help visitors to visualise the sizes of different groups of people by using one grain of rice per person. For example, one pile of rice might represent the number of millionaires in the world, while another might represent the number of refugees. The piles of rice representing different groups of people are juxtaposed in ways that are variously thought-provoking, shocking or amusing. Figure 23 overleaf shows part of one version of the installation.

For small groups of people, such as the number of people who have walked on the moon, the rice can be counted, but for large groups, such as the populations of countries, the rice has to be weighed. So the people setting out the piles of rice need a means of calculating the weight of rice required for a given number of people. One way to do this is to use a formula. Since the number of people represented and the weight of rice needed are directly proportional, the formula will be of the type that you have seen in this subsection. The next example shows how a suitable formula can be found.

All the rice used in each installation is later returned to be used for food.



Figure 23 Part of the art installation 'Of all the people in all the world' by the Birmingham-based theatre company *Stan's Cafe*

Example 8 Finding a constant of proportionality for the rice

The publicity information for 'Of all the people in all the world' states that 1 kg of rice is needed to represent 60 000 people.

- Find a formula for the weight of rice needed in terms of the number of people represented. Use the letters r and p to represent the weight of rice in kilograms and the number of thousands of people represented, respectively.
- Use the formula to find the weight of rice needed to represent the number of students of The Open University in 2006–07, which was about 224 000.

Solution

- Since r and p are directly proportional, the formula expressing their relationship is of the form

$$r = kp,$$

where k is a constant. Also, $r = 1$ when $p = 60$. Substituting these values into the formula gives

$$1 = k \times 60.$$

We now solve this equation to find the value of k .

$$\text{Divide by 60: } \frac{1}{60} = k$$

Substituting $k = \frac{1}{60}$ into the equation $r = kp$ gives

$$r = \frac{1}{60}p.$$

This is the required formula for r in terms of p .

- (b) The number of OU students in 2006–07 was 224 000, which gives $p = 224$, since p is measured in thousands. Substituting this value of p into the formula above gives

$$r = \frac{1}{60} \times 224 = 3.73 \text{ (to 3 s.f.)}.$$

So approximately 3.73 kg of rice is needed.

Here is a similar example for you to try.

Activity 13 Finding a constant of proportionality

Epilepsy in children is sometimes treated with anti-epileptic drugs. The amount of an anti-epileptic drug needed by a child each day is directly proportional to the child's weight. For a particular drug, the daily dose for a child of weight 20 kg is 400 mg.

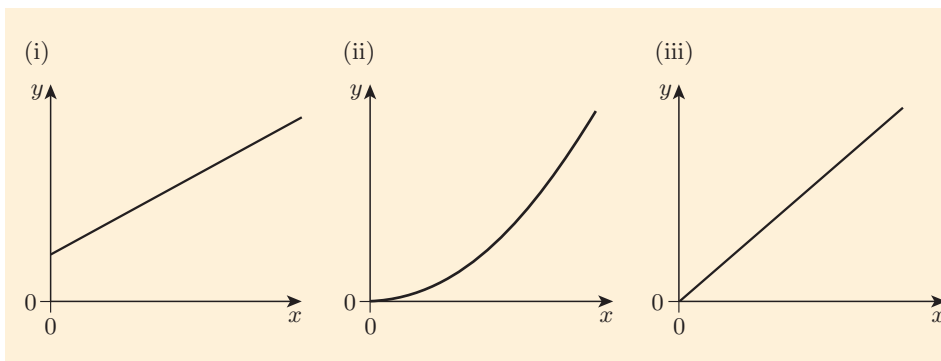
- (a) Find a formula for the amount of this drug needed each day, in terms of the weight of the child treated. Use d to represent the weight in milligrams of the drug needed each day, and c to represent the weight of the child in kilograms.
- (b) How much of the drug is needed each day for a child of weight 35 kg?

You have seen that if two quantities x and y are directly proportional, then their relationship is expressed by an equation of the form $y = kx$, where k is a non-zero constant. So the graph illustrating the relationship is a straight line through the origin with gradient k .

This means that if you have a graph of the relationship between two quantities, then you can immediately recognise whether the quantities are directly proportional. If the graph is a straight line through the origin, then the quantities are directly proportional and the constant of proportionality is the gradient of the graph. If the graph has any other form, then the quantities are not directly proportional.

Activity 14 Recognising graphs of direct proportion relationships

- (a) Which of the following graphs represent direct proportion relationships?



- (b) Look back at graph (a) in Example 6 on page 83. Is temperature in degrees Fahrenheit directly proportional to temperature in degrees Celsius?

3.3 The general equation of a straight line

In this subsection you will see that nearly every straight line that you can draw on a graph is described by an equation of the form

$$y = mx + c,$$

where m and c are constants. (The only exceptions are vertical lines.) In the following activity you are asked to use Graphplotter to investigate the graphs of equations of this form.



Graphplotter

Activity 15 Investigating graphs of equations of the form $y = mx + c$

Use Graphplotter, with the equation $y = mx + c$ selected.

- Click the 'Options' tab, then click ' y -intercept' to display the coordinates of the point where the graph crosses the y -axis. Click the 'Functions' tab to return to the main panel.
- Set $m = 2$ and $c = 0$, and check that you obtain the line through the origin with gradient 2, as expected. (To set m and c to these values, type them into the boxes and press 'Enter'.)
- Now set $c = 1$ and observe the effect on the graph. What is the y -intercept of the new graph?
- Choose some other values of c (both positive and negative) and repeat part (c) for each of these values.
- What is the same and what is different about all the graphs in parts (b), (c) and (d)?
- Use the sliders to experiment with changing the values of m and c , and observe the effect on the graph.

Here is an explanation of what you observed in Activity 15. Figure 24(a) shows the line that passes through the origin and has gradient 2. As you saw earlier in this section, the y -coordinate of every point on this line is twice the x -coordinate, so the line has equation $y = 2x$.

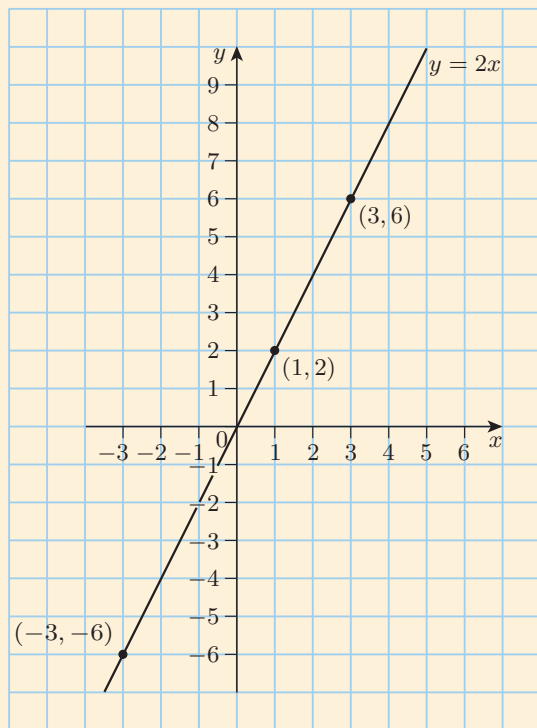
Now consider the equation

$$y = 2x + 1.$$

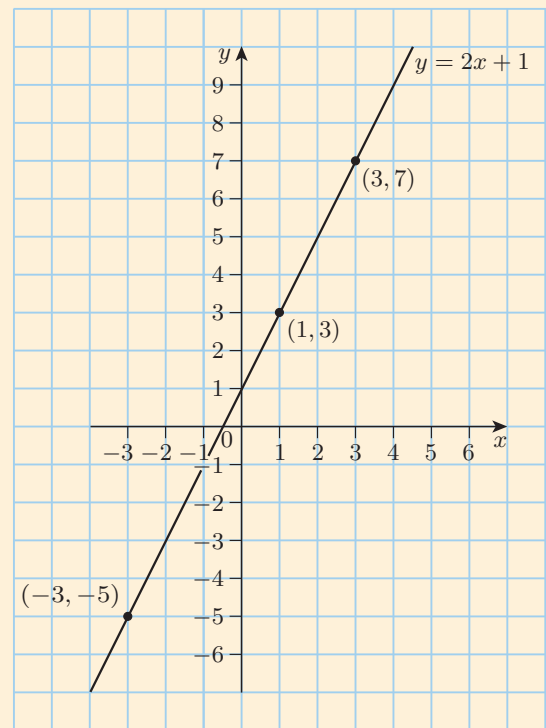
To obtain the graph of this equation, you could take every point on the graph of the equation $y = 2x$, and add 1 to the y -coordinate. This has the effect of moving the graph of $y = 2x$ up the y -axis by 1 unit. The result is the line shown in Figure 24(b). It has the same gradient as the line $y = 2x$, but it does not pass through the origin – instead, its y -intercept is 1.

A similar argument can be applied to any equation of the form $y = mx + c$. The graph of the equation $y = mx + c$ is obtained by moving the graph of the equation $y = mx$ along the y -axis by c units. You know that the graph of $y = mx$ is the straight line with gradient m through the origin, so the graph of $y = mx + c$ is the straight line with gradient m and y -intercept c . This applies whether c is positive or negative.

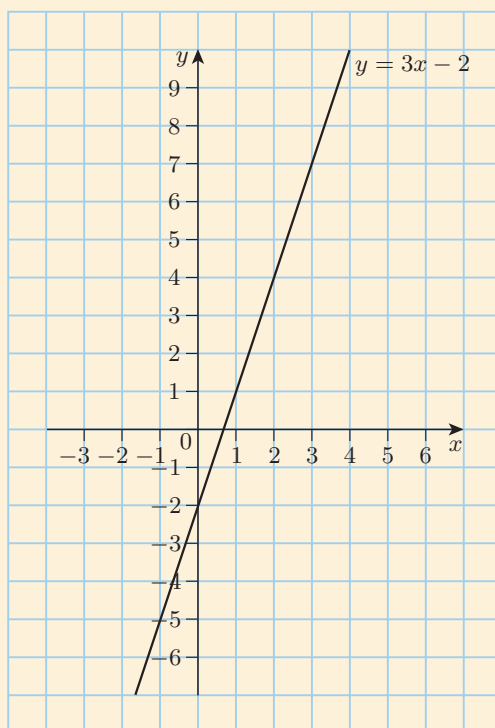
Figure 25 shows two more examples. Notice that the line $y = 3x - 2$ has gradient 3 and y -intercept -2 , and the line $y = -x + 3$ has gradient -1 and y -intercept 3.



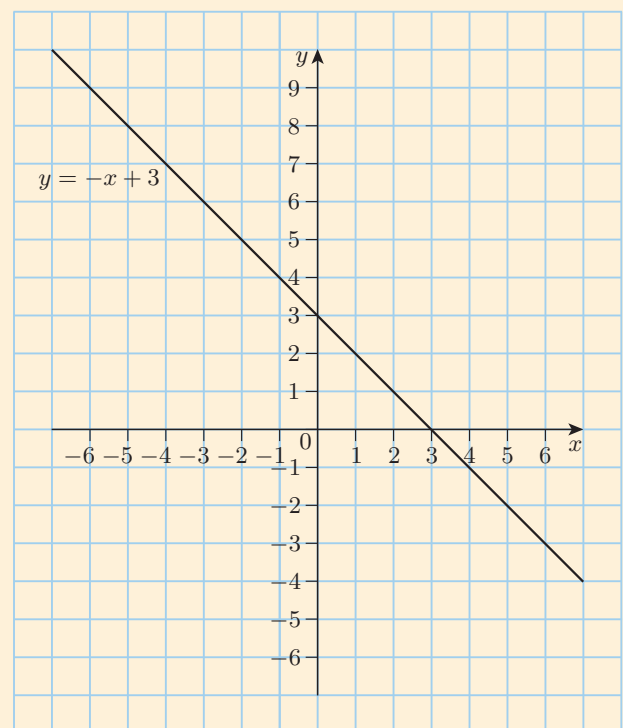
(a)



(b)

Figure 24 The lines (a) $y = 2x$ and (b) $y = 2x + 1$ 

(a)



(b)

Figure 25 The lines (a) $y = 3x - 2$ and (b) $y = -x + 3$

The new fact that you have met is summarised below.

The general equation of a straight line

The line with gradient m and y -intercept c has equation

$$y = mx + c.$$

Example 9 Writing down gradients, y -intercepts and equations of lines

- (a) Write down the gradient and y -intercept of the line $y = 4x - 3$.
 (b) Write down the equation of the straight line with gradient -5 and y -intercept 2 .

Solution

- (a) Compare the equation with the equation $y = mx + c$ to identify the values of the gradient m and the y -intercept c . Remember that $4x - 3$ is the same as $4x + (-3)$.

The coefficient of x is 4 , so the gradient is 4 .

The constant term is -3 , so the y -intercept is -3 .

- (b) The equation is $y = mx + c$, where m is the gradient and c is the y -intercept.

The equation is $y = -5x + 2$.

Activity 16 Writing down gradients, y -intercepts and equations of lines

- (a) Write down the gradients and y -intercepts of the following lines.
 (i) $y = 2x - 1$ (ii) $y = -3x + 4$ (iii) $y = \frac{x}{5} - \frac{2}{5}$
 (b) Write down the equations of the following lines.
 (i) The line with gradient 4 and y -intercept -10
 (ii) The line with gradient -1 and y -intercept 5
 (iii) The line with gradient 0 and y -intercept 3
 (c) For each of the lines in part (b), write down the equation of the line that is parallel to the given line and has y -intercept 2 . (Remember that two lines are parallel if they have the same gradient.)

The next example shows you how to use the equation of a line to find its x -intercept.

Example 10 Finding an x -intercept from an equation

Find the x -intercept of the line with equation $y = 4x - 3$.

Solution

The x -intercept is the value of x when $y = 0$.

Putting $y = 0$ gives

$$4x - 3 = 0.$$

We now solve this equation.

$$\text{Add 3:} \quad 4x = 3$$

$$\text{Divide by 4:} \quad x = \frac{3}{4}$$

Hence the x -intercept is $\frac{3}{4}$.

Remember that the *solution* of an equation is the value of the unknown for which the equation is true.

Activity 17 Finding x -intercepts from equations

Find the x -intercepts of the following lines.

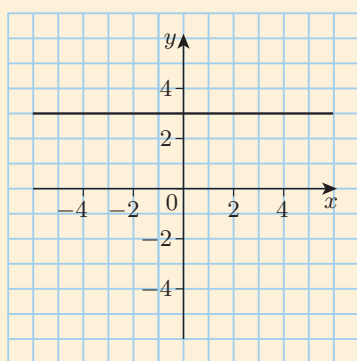
(a) $y = 2x - 1$ (b) $y = -3x + 4$ (c) $y = \frac{x}{5} - \frac{2}{5}$

(These are the lines from Activity 16(a).)

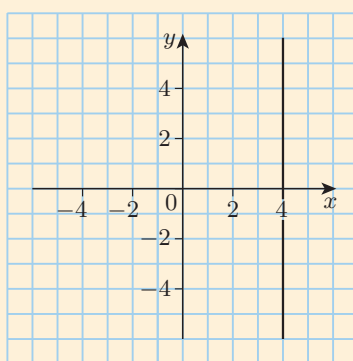
The equation of a horizontal or vertical line

Figure 26(a) shows the line with gradient 0 and y -intercept 3. In Activity 16(b)(iii) you saw that the equation of this line is $y = 3$. This equation is found by substituting $m = 0$ and $c = 3$ into the general equation $y = mx + c$.

An alternative way to find the equation of this line is as follows. Every point on this line has y -coordinate 3, so the equation $y = 3$ describes each point on the line and is therefore the equation of the line.



(a)



(b)

Figure 26 Graphs showing (a) a horizontal line and (b) a vertical line

What about vertical lines? You cannot use the fact that the line with gradient m and y -intercept c has equation $y = mx + c$ to find the equation of a vertical line, because the gradient of a vertical line is undefined.

Vertical lines are the only lines that do not have equations of the form $y = mx + c$.

However, you can find an equation for a vertical line by using an approach similar to the alternative approach suggested above for horizontal lines.

For example, consider the vertical line that has x -intercept 4, which is shown in Figure 26(b) on the previous page. Every point on this line has x -coordinate 4, so the equation $x = 4$ describes each point on the line and is therefore the equation of the line.

Equations of horizontal and vertical lines

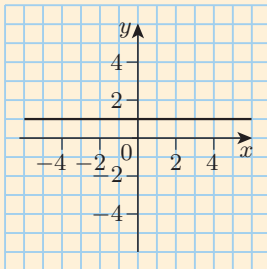
The horizontal line with y -intercept a has equation $y = a$.

The vertical line with x -intercept a has equation $x = a$.

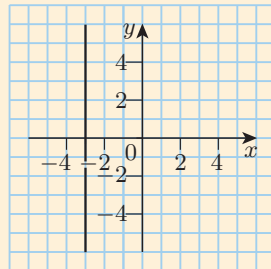
Activity 18 Writing down equations of horizontal and vertical lines

(a) Write down the equations of the lines shown below.

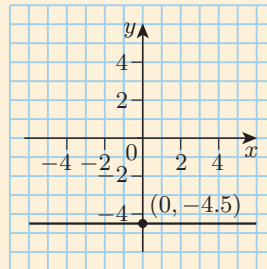
(i)



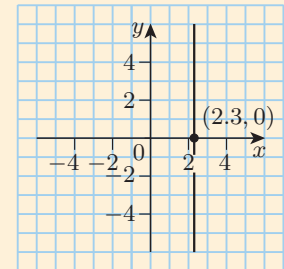
(ii)



(iii)



(iv)



(b) Write down the equations of the x - and y -axes.

3.4 Drawing lines from their equations

When you are working on a question that involves the equation of a straight line, it often helps to draw the line on a graph. One way to do this is to use the following strategy. This strategy is particularly useful when you just want a quick sketch rather than a very accurate graph.

Strategy To draw the line $y = mx + c$ (gradient method)

Use the values of m and c .

1. Mark the point $(0, c)$ that corresponds to the y -intercept.
2. Count 1 unit right and m units up from this point, and mark the point that you reach. (If m is negative, then you count down rather than up.)
3. Draw the straight line through the two points.

(If the value of m is small, then in step 2 it might be easier to count, say, 2 units right and $2m$ units up, or 3 units right and $3m$ units up, and so on – choose a convenient multiple.)

This strategy is demonstrated in the next example.

Example 11 Drawing a line from its equation (gradient method)

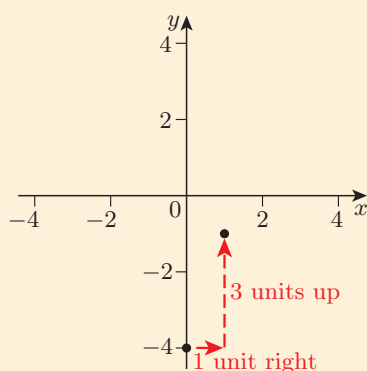
Tutorial clip

Draw the graph of the equation $y = 3x - 4$.

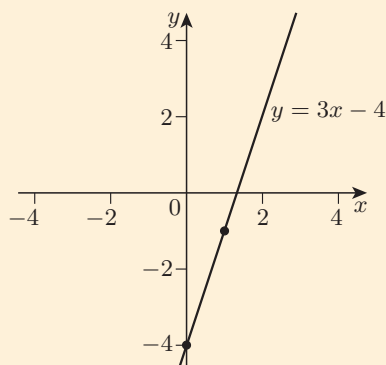
Solution

This equation is of the form $y = mx + c$, with $m = 3$ and $c = -4$. So the graph is a straight line with gradient 3 and y -intercept -4 .

Draw the axes and mark the scales. The y -intercept is -4 , so mark the point $(0, -4)$. The gradient of the line is 3, so count 1 unit right and 3 units up, and mark the point that you get to.



Draw the straight line through the two points and label it with the equation.



Once you have drawn a straight line from its equation, it is a good idea to check that it looks roughly as you would expect.

One check that you can do is to consider whether the slope looks reasonable. For example, you know from its equation that the line $y = 3x - 4$ has gradient 3. This is positive, so you would expect the line drawn in Example 11 to slope up, and it does. Also, this gradient is greater than 1 in size, so if the axis scales are equal then you would expect the line to make an angle of more than 45° with the x -axis – again, it does.

As a further check, you can work out a third point on the line. For example, substituting $x = 2$ into the equation gives

$$y = 3 \times 2 - 4 = 6 - 4 = 2.$$

So the point $(2, 2)$ should lie on the line, which it does.

Remember from Unit 2 that when you plot a graph to illustrate *data*, you should provide a title and the source of the data.

There is usually no need to provide a title for a graph that illustrates an equation. Instead, you should label the line (or curve) with the equation, as shown in Example 11.

Here is an alternative strategy for drawing a straight line from its equation. This strategy can take a little longer, but it is often preferable when accuracy is important.

Strategy To draw the line $y = mx + c$ from its equation (two-point method)

1. Find the coordinates of two points on the line, by choosing two values of x and substituting them into the equation to find the corresponding values of y .
2. Plot the points on a graph, and draw the straight line through them.

When you use this strategy, you should also find and plot a third point on the line, as a check. The strategy is illustrated in the next example.



Tutorial clip

Example 12 Drawing a line from its equation (two-point method)

Draw the graph of the equation $y = -\frac{3}{4}x + 1$.

Solution

The equation is of the form $y = mx + c$, so the graph is a straight line.

☞ Choose two values of x and calculate the corresponding values of y . Try to choose values of x that make the calculation simple. ☞

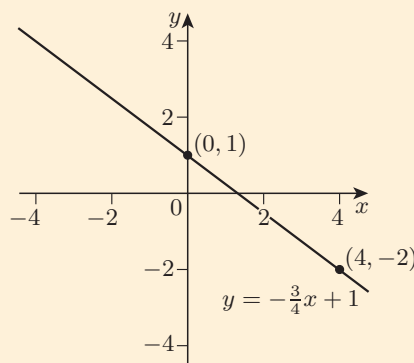
When $x = 0$, $y = -\frac{3}{4} \times 0 + 1 = 1$.

So the point $(0, 1)$ lies on the line.

When $x = 4$, $y = -\frac{3}{4} \times 4 + 1 = -3 + 1 = -2$.

So the point $(4, -2)$ also lies on the line.

☞ Plot the points and draw the straight line through them. Label the line with its equation. ☞



(Check: When $x = -4$, $y = -\frac{3}{4} \times (-4) + 1 = 3 + 1 = 4$. So the point $(-4, 4)$ should also lie on the line, which it does.)

An alternative way to find the point $(0, 1)$ is to observe from the equation that the y -intercept is 1.

The strategies above do not apply to vertical lines. However, you have seen that a vertical line has an equation of the form $x = a$, where a is the x -intercept, and a horizontal line has an equation of the form $y = a$, where a is the y -intercept. You should be able to recognise equations of these forms and hence draw their graphs.

Activity 19 Drawing lines from their equations

Draw the graphs of the following equations. You can use any method that you have seen in this subsection.

- (a) $y = 2x + 3$ (b) $y = -2x + 4$ (c) $y = 3$
 (d) $y = \frac{1}{2}x - 1$ (e) $x = \frac{5}{2}$

In this unit you have seen two different ways of producing a graph of an equation. If you are asked to *plot* a graph of an equation, then you should construct a table of values, plot the points and draw the line or curve through them in the way that you saw in Subsection 1.2. On the other hand, if you are asked to *draw* or *sketch* the graph of an equation, then you should do so using facts that you can deduce from the equation, such as the gradient and y -intercept of a straight line. You have seen how to do this for equations of the form $y = mx + c$ in this subsection, and you will find out how to draw graphs of other types of equation later in the module.

If you are drawing a sketch of a graph for your own use, for example to help you to check an equation of a line, then you can include as much or as little detail as you like. However, if a question asks you for a sketch of a graph, then you must include all the relevant detail, such as the values of the intercepts, and the equation of the line or curve.

3.5 Finding the equations of lines

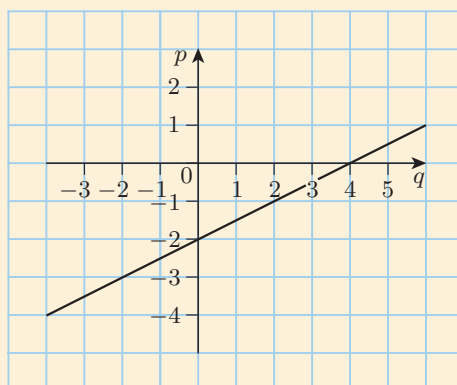
You have seen that if you know the gradient and y -intercept of a line, then you can immediately write down the equation of the line.

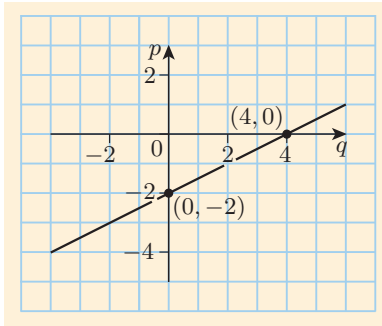
Sometimes you might have different information about a line, and want to determine its equation. For example, you might have a graph of the line, or you might know the coordinates of two points on the line.

In the next example, a graph of a line is used to find the gradient and the y -intercept, so that the equation can be written down. This involves using several of the skills that you developed earlier. In this example the variables on the axes are p and q instead of the usual x and y .

Example 13 Finding the equation of a line from a graph

Find the equation of the line shown below.



**Solution**

Work out the gradient. To do this, choose two points on the line with coordinates that are easy to read.

The points $(0, -2)$ and $(4, 0)$ lie on the line, as shown in the margin. So the gradient is

$$\frac{0 - (-2)}{4 - 0} = \frac{2}{4} = \frac{1}{2}.$$

Read off the vertical intercept from the graph.

The vertical intercept is -2 .

Substitute the gradient and vertical intercept into the general equation of a line. The variables here are not x and y , so in the general equation $y = mx + c$ replace x by the variable on the horizontal axis, q , and y by the variable on the vertical axis, p .

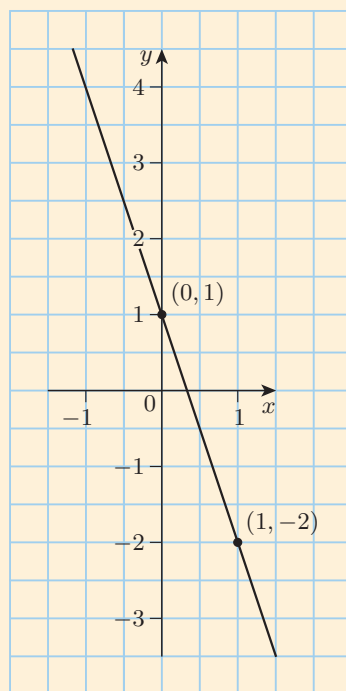
The equation of the line is

$$p = \frac{1}{2}q - 2.$$

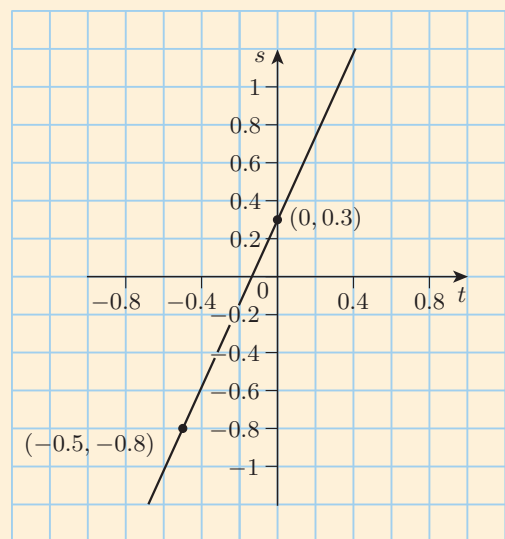
When you work out the equation of a line from its graph, remember that if the exact coordinates of points on the graph are given, then you should use these to work out the equation, rather than reading off further coordinates from the graph, as your readings may not be accurate.

Activity 20 Finding the equation of a line from a graph

Find the equations of the lines shown below.



(a)



(b)

Sometimes the information that you have about a line does not include the vertical intercept. For example, you might have a graph showing a part of a line that is not near the y -axis, or you might know only the coordinates of two points on the line. In cases like these you can usually find the equation of the line by using the gradient and the coordinates of any point on the line. The method is illustrated in the next example.

When you are trying to find the equation of a line, and you do not have a graph already, it is a good idea to begin by sketching the line. Then you can use your sketch to check your answer.

Example 14 Finding the equation of a line from the gradient and a point

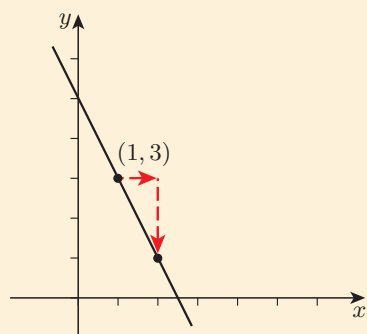


Tutorial clip

Find the equation of the line that has gradient -2 and passes through the point $(1, 3)$.

Solution

Sketch the line – you can use the gradient method (but starting at the given point on the line rather than the y -intercept). As the sketch is just to help you to check your answer, you can save time by omitting some of the detail, such as the numbers on the axes.



Substitute the gradient into the general equation of a line.

The equation of the line is $y = mx + c$, where m is the gradient and c is the y -intercept. The gradient is -2 , so the equation is

$$y = -2x + c.$$

Substitute the coordinates of a point on the line into the equation and solve the resulting equation to find the value of the y -intercept c .

The point $(1, 3)$ lies on the line so its coordinates satisfy the equation. Substituting $x = 1$ and $y = 3$ into the equation gives

$$3 = -2 \times 1 + c; \quad \text{that is,} \quad 3 = -2 + c.$$

Adding 2 to both sides gives

$$5 = c.$$

Now the values of both m and c are known, so write down the equation of the line.

The equation of the line is

$$y = -2x + 5.$$

(Check: This is the equation of a line with y -intercept 5, and the y -intercept on the sketch does appear to be 5.)

This strategy applies to *non-vertical* lines.

Here is a summary of the strategy used in Example 14.

Strategy To find the equation of a line when you do not know the vertical intercept

1. Find the gradient m of the line and substitute it into the general equation $y = mx + c$.
2. Substitute the coordinates of a point on the line into the equation of the line from step 1, and solve the resulting equation to find the value of the y -intercept c .
3. Use the values of m and c to write down the equation of the line.

Here is a similar activity for you to try.

Activity 21 Finding the equation of a line from the gradient and a point

Find the equation of the line that has gradient 3 and passes through the point $(-2, -3)$.

You can use the strategy above to find the equation of a straight line if the only information that you have is the coordinates of two points on the line. This is illustrated in the next example.



Tutorial clip

Example 15 Finding the equation of a line through two points

Find the equation of the line that passes through the points $(1, 2)$ and $(3, 5)$.

Solution

A sketch of the line is shown in the margin.

The gradient of the line is

$$\frac{5 - 2}{3 - 1} = \frac{3}{2}.$$

So the equation of the line is $y = \frac{3}{2}x + c$, where c is a constant.

Also, the line passes through the point $(1, 2)$. Substituting these coordinates into the equation gives

$$2 = \frac{3}{2} \times 1 + c; \quad \text{that is,} \quad 2 = \frac{3}{2} + c.$$

Subtracting $\frac{3}{2}$ from both sides gives

$$\frac{1}{2} = c.$$

So the equation of the line is

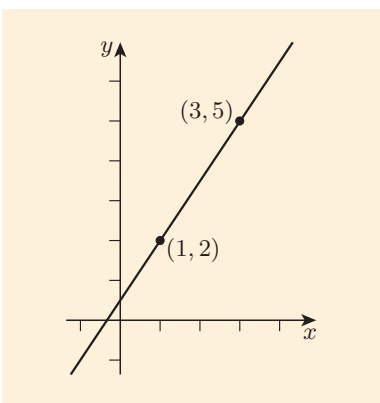
$$y = \frac{3}{2}x + \frac{1}{2}.$$

Check that the coordinates of the other point satisfy the equation.

(Check: Substituting $x = 3$ into the equation gives

$$y = \frac{3}{2} \times 3 + \frac{1}{2} = \frac{9}{2} + \frac{1}{2} = \frac{10}{2} = 5,$$

so the coordinates $(3, 5)$ satisfy the equation.)



Activity 22 Finding the equation of a line through two points

- (a) Plot the points $(1, 4)$ and $(3, -2)$ on a graph, and draw the straight line through them.
- (b) Find the equation of the line.

In the final example in this subsection, the equations of some horizontal and vertical lines are found from given information.

Example 16 Finding the equations of horizontal and vertical lines

Write down the equations of the following straight lines.

- (a) The line that is parallel to the x -axis and passes through the point $(0, 2)$
- (b) The line that is parallel to the y -axis and passes through the point $(-3, 1)$

Solution

Sketches of the lines are shown in the margin.

- (a)  A line parallel to the x -axis has an equation of the form $y = a$. 

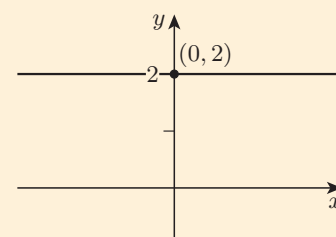
The equation is $y = a$, where a is a constant. Since the point $(0, 2)$ satisfies the equation, the equation is $y = 2$.

- (b)  A line parallel to the y -axis has an equation of the form $x = a$. 

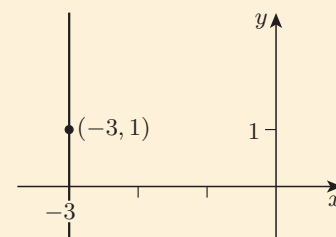
The equation is $x = a$, where a is a constant. Since the point $(-3, 1)$ satisfies the equation, the equation is $x = -3$.



Tutorial clip



(a)



(b)

Activity 23 Finding the equations of horizontal and vertical lines

Write down the equations of the following straight lines.

- (a) The line that is parallel to the y -axis and passes through the point $(1, 0)$
- (b) The line that is parallel to the x -axis and passes through the point $(2, -4)$

In this section you have seen that an equation of the form $y = mx + c$, where m and c are constants, represents a straight line graph with gradient m and y -intercept c , and that every line (except vertical ones) can be represented by an equation of this form. This is an important result, as many situations can be modelled by using a straight line graph or a linear equation, and it can be helpful to work with both the graph and the equation. You will see some examples of this in the next section and in Unit 7.

The result stated in the paragraph above is the reason why an equation in x , each of whose terms is either a constant term or a number times x (after expanding any fractions or brackets in the equation), is called a **linear equation**. In other words, a **linear equation** in x is an equation of

the form $mx + c = 0$, where m and c are constants with $m \neq 0$, or an equation that can be rearranged into this form. Similarly, any expression of the form $mx + c$ where m and c are constants with $m \neq 0$ is called a **linear expression** in x . A function with a rule of the form $y = mx + c$ where m and c are constants with $m \neq 0$ is called a **linear function**.

4 Linear models from data

In Subsection 1.3 you saw that it is sometimes possible to model a set of paired data using a straight line. In this section you will see how to find the equation of the best line to fit a set of data, and how to measure how good a fit the line is. You will see some real-life datasets modelled in this way.

4.1 Regression lines

At the time of writing, the tallest man to have had his height officially recorded was Robert Pershing Wadlow, who was born on 22 February 1918 in Alton, Illinois, USA. Mr Wadlow is pictured in Figure 27, and Table 2 gives his height in centimetres measured at various ages.

Table 2 The height of Robert Wadlow at various ages

Age (years)	5	8	9	10	12	14	16	18	20	21	22
Height (cm)	163	183	188	196	211	226	239	254	262	264	272

Source: www.altonweb.com

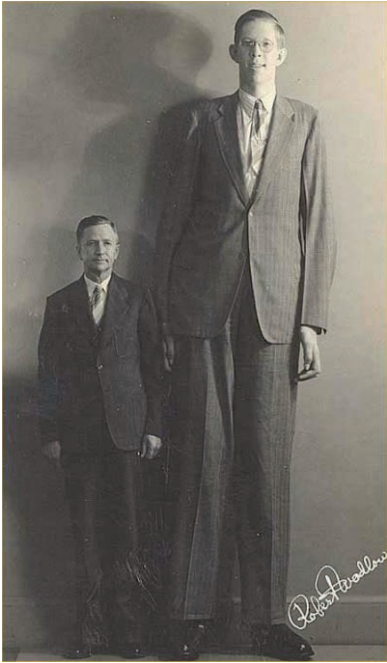


Figure 27 Robert Wadlow with his father, Harold Franklin Wadlow

It is difficult to tell how quickly Mr Wadlow grew over the years from the table alone. A better impression can be obtained by plotting the data in the table on a scatterplot, as shown in Figure 28.

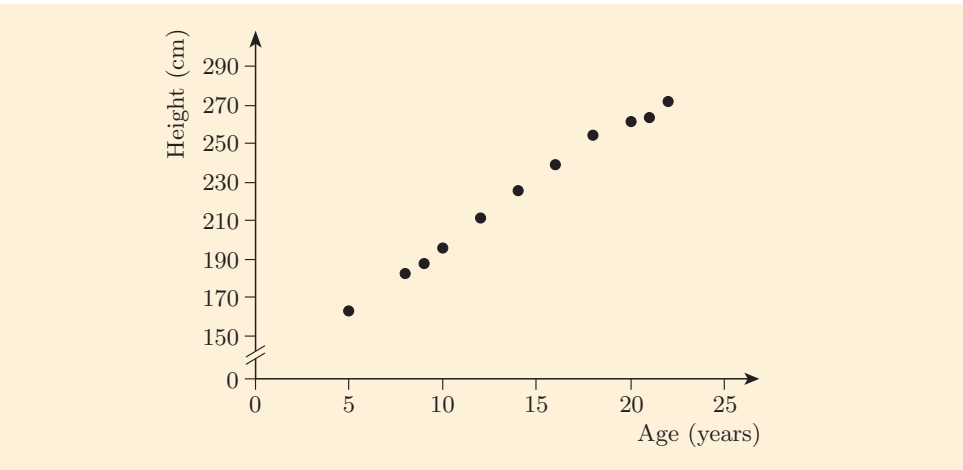


Figure 28 The height of Robert Wadlow plotted against his age

It looks as if the points in Figure 28 lie approximately in a straight line. However, it is not possible to draw a straight line that passes exactly through all the points. This may be because Mr Wadlow did not grow at a steady rate. Alternatively, it could be a consequence of how the data were collected – for example, Table 2 does not give information on the time of year when each measurement was taken. If this information were available, then a measurement taken in January, say, would be plotted at a slightly different position to a measurement taken in December, and the points might lie more closely in a straight line.

So the data on Mr Wadlow's height can be modelled approximately, but not exactly, by a straight-line graph. But what is the 'best' straight line for the data points plotted? For example, the line drawn in Figure 29(a) clearly fits the data better than the line drawn in Figure 29(b), but it may still not be the *best* line.

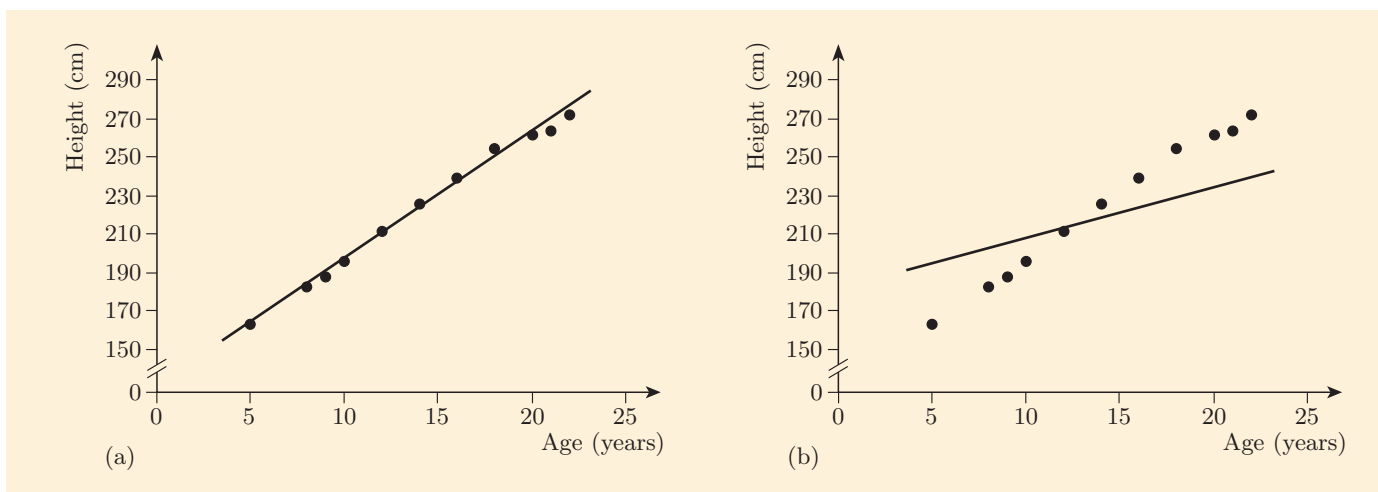


Figure 29 Straight lines drawn on the graph of Mr Wadlow's height

Roughly, a line is a good fit to data like these if the distances from the points to the line are small. Statisticians have found that the best way to measure how well a line fits the data is as follows. The vertical distances from the points to the line are measured, as illustrated in Figure 30, then each of these distances is squared, and finally the squared distances are all added up. The smaller the sum of the squared distances, the better the fit of the line. There is always just one line for which this sum is the smallest, and it is known by various names: the *regression line*, the *least squares fit line*, the *best fit line* or the *trend line*. In this module the first of these terms is used, the **regression line**.

You can learn why this is the best way to measure the fit of a line, and how this method can be turned into a formula for the equation of the best line, if you go on to study modules on statistics.

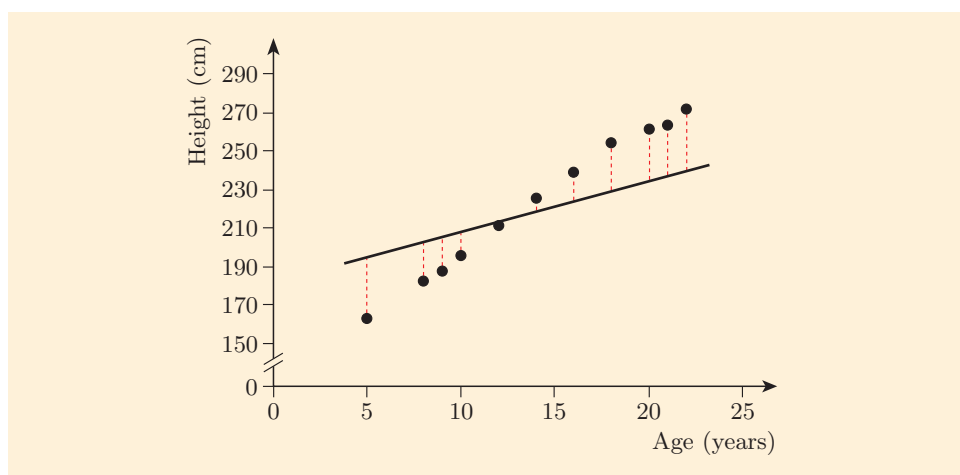


Figure 30 The distances of the points from a line

There is a formula for calculating the equation of the regression line from the data values, but it is usually easiest to use software to do the calculation. The Dataplotter software supplied with the module can calculate the equation of a regression line, as you will see in the next activity. There are instructions for using Dataplotter in the MU123 Guide.



Dataplotter

Dataplotter calculates the numbers in a regression equation to five significant figures, but it does not display trailing zeros. This means that, for example, the coefficient 6.5140 in the regression equation found in Activity 24 is displayed as 6.514.

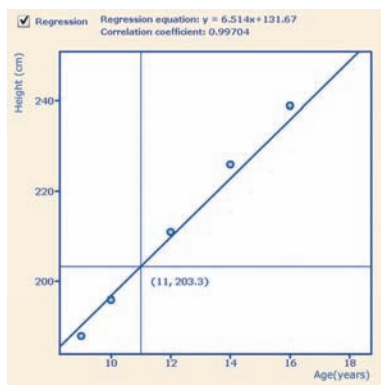


Figure 31 Using Dataplotter to obtain an estimate for Mr Wadlow's height when he was 11



Dataplotter

You can move the graph by dragging it or using the arrow buttons.

Activity 24 Finding a regression line

- Open Dataplotter and choose the Scatterplot tab. Use the drop-down lists to choose the datasets '# Wadlow age' and '# Wadlow ht' for the first and second column, respectively, to obtain a scatterplot.
- Click the 'Regression' box above the scatterplot to add the regression line to the scatterplot. Write down the equation of the regression line, which is displayed at the top of the scatterplot.

In Activity 24 you should have found that the equation of the regression line for the data on Robert Wadlow's growth is

$$y = 6.5140x + 131.67,$$

where y is the height in cm and x is the age in years.

The regression line on the graph and its equation are two different ways of describing the same model for Mr Wadlow's growth, and they can both be used to make estimates about his height at different ages. For example, if you want to estimate Mr Wadlow's height when he was 11 years old, then you can read off the appropriate value from the graph. One way to do this is to move the cursor over the Dataplotter scatterplot and regression line produced in Activity 24, to find the y -coordinate of the point on the line with x -coordinate 11. This is shown in Figure 31 – zooming in increases the precision. This method gives a height of about 203 cm. Alternatively, you can substitute $x = 11$ into the equation of the regression line, which gives

$$y = 6.5140 \times 11 + 131.67 \approx 203.$$

So an estimate for Mr Wadlow's height when he was 11 is approximately 203 cm. This estimate is rounded to three significant figures because this is the level of precision of the height measurements in the data.

When you use a model to estimate a value in this way, it is important to consider how accurate the estimate is likely to be. The regression line found above for Mr Wadlow's growth is based on his height at ages from 5 to 22, and age 11 is within this range. Also, the regression line is very close to all the points plotted. So the estimate given by the model for Mr Wadlow's height at age 11 is probably reasonably accurate.

You can often use a model based on data to estimate values that are not given in the data. If the estimated value is within the *range* of the data, then this process is known as **interpolation**. As long as the model fits the data reasonably well, interpolation can provide reasonable estimates.

Activity 25 Using a regression line

- Use the equation of the regression line for Mr Wadlow's growth to estimate his height at the age of 24 years.
- Use the Dataplotter graph found in Activity 24 to check your answer to part (a). You will have to move the graph within the window.
- Why might the estimate found in parts (a) and (b) be unreliable?

In Activity 25 you used a model based on a dataset to estimate a value that lies outside the range of the dataset. This is called **extrapolation**,

and it can be unreliable. For example, if you read off the y -intercept of the line that models Mr Wadlow's growth (or substitute $x = 0$ into the equation of the line), then you will find that the model predicts that Mr Wadlow's height at age 0 was about 130 cm. This sounds very unlikely! In fact, although Mr Wadlow's height was not recorded at birth, his weight was a healthy 3.8 kg and his rapid growth started soon after birth.

In general it is unwise to extrapolate too far from the range of a dataset, unless you are confident that the conditions that apply to the values in the dataset also apply to values outside its range. (Sometimes there may be independent scientific evidence to suggest that this is the case.)

The gradient of a line that models a relationship tells you the rate of change of the quantity on the vertical axis with respect to the quantity on the horizontal axis. For example, the gradient of the line that models Mr Wadlow's growth is 6.5140 cm/year. So on average Mr Wadlow grew about 6 to 7 cm each year between the ages of 5 and 21.

Because of inevitable dirt and imperfections in the disc, an audio CD player cannot successfully read every single 'bit' of data from a CD. Sophisticated mathematical error-correction techniques usually allow missing data to be reconstructed, but when too much data has been lost, the CD player may have to resort to interpolation to guess what the missing data might be.

4.2 Correlation coefficients

It is important to know how accurate a prediction provided by a regression line is likely to be, so it is useful to have some indication of how well a regression line fits its data points. To provide this, statisticians have developed a measure known as the **correlation coefficient**, which is a number calculated from the data pairs, and is often denoted by r . The value of the correlation coefficient indicates how well the regression line fits the data pairs.

There is a formula for the correlation coefficient, but, as with regression lines, it is easiest to use a calculator or computer to calculate it. The software Dataplotter supplied with the module can calculate correlation coefficients, as you will see in the next activity. In this activity you will explore how the correlation coefficient varies for different datasets. After the activity, you will see how the value of the correlation coefficient is interpreted.

You may see the formula for the correlation coefficient if you go on to study modules on statistics.

Activity 26 Exploring the correlation coefficient



Dataplotter

Use Dataplotter, with the Scatterplot tab selected. If any datasets are already selected, then first remove each of them by clicking on the 'New' button to produce a new empty dataset.

- Plot the four points (0, 0), (1, 2), (3, 6) and (4, 8) on the scatterplot, by entering the x -values 0, 1, 3 and 4 in the first column, and the y -values 0, 2, 6 and 8 in the second column. These points all lie in a straight line with a positive gradient.
- If the Regression box is not already ticked, then click it to display the regression line. The correlation coefficient is displayed at the top of the scatterplot, under the equation of the regression line. What is its value?
- Now plot some more points that do *not* lie on the regression line. You can do this by holding down the shift key and clicking on the graph. What happens to the regression line and the correlation coefficient?
- Clear each of the columns by clicking on 'Clear'. Then repeat parts (a), (b) and (c), but this time plot four points that lie on a line with a *negative* gradient, such as (0, 0), (1, -2), (3, -6) and (4, -8).

The correlation coefficient described here was invented by Francis Galton (1822–1911). As well as pursuing his statistical work, Galton founded the modern system of weather-mapping (inventing the word ‘anticyclone’), played a major part in the development of fingerprint classification, and invented the dog whistle, which dogs can hear but people cannot.

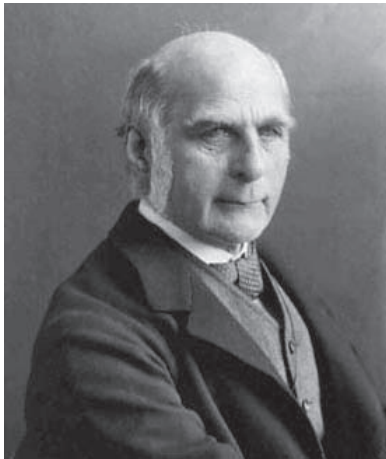


Figure 33 Francis Galton (1822–1911) was a pioneer in the use of regression techniques

The correlation coefficient of a set of paired data is always between -1 and 1 , inclusive. If the correlation coefficient is *positive*, then the regression line for the data has a positive gradient, which means that one of the quantities tends to increase as the other increases. In this case, the quantities are said to have a **positive correlation**. On the other hand, if the correlation coefficient is *negative*, then the regression line has a negative gradient, which means that one of the quantities tends to decrease as the other increases. In this case, the quantities are said to have a **negative correlation**.

The correlation coefficient is exactly 1 or exactly -1 when all the data points lie exactly on the regression line. When this happens, the quantities are said to have a **perfect correlation**, as illustrated in Figure 32.

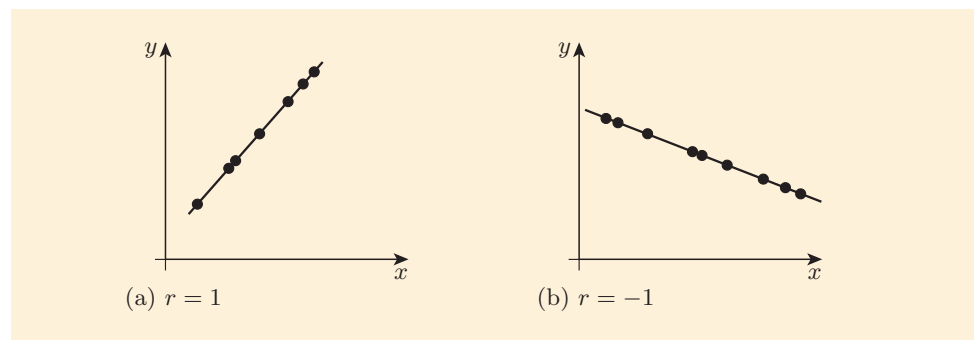


Figure 32 (a) Perfect positive correlation ($r = 1$).
(b) Perfect negative correlation ($r = -1$).

If the correlation coefficient is close to 1 or -1 , then the data points lie close to the regression line, as illustrated in Figure 34, and the quantities are said to have a strong correlation.

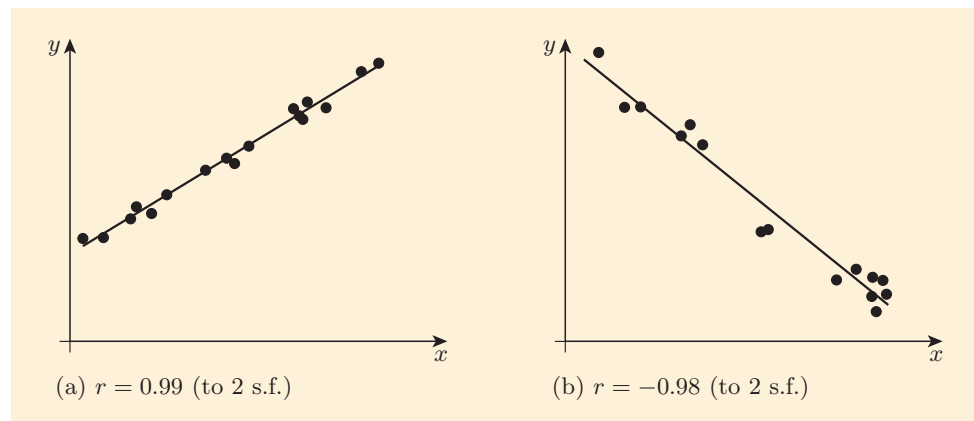


Figure 34 (a) Strong positive correlation (r close to 1).
(b) Strong negative correlation (r close to -1).

If the correlation coefficient is closer to zero, then the data points are scattered further from the regression line, as illustrated in Figure 35. In general, the closer the correlation coefficient is to zero, the weaker is the correlation between the quantities. Graphs (a) and (b) in Figure 35 illustrate **weaker correlation** than the graphs in Figure 34. In both cases, the correlation coefficient, r , is closer to zero than it is for the strong correlations shown in Figure 34. When the correlation coefficient is zero there is no correlation, as shown in graph (c).

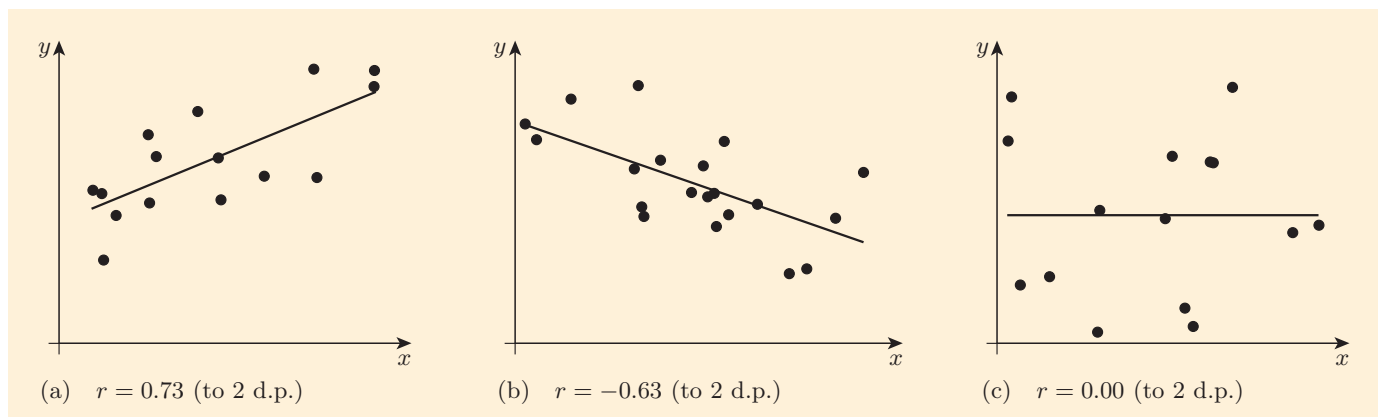


Figure 35 (a) Weaker positive correlation (r positive but closer to 0).
 (b) Weaker negative correlation (r negative but closer to 0). (c) Zero correlation ($r = 0$).

The key facts about correlation coefficients are summarised below.

Correlation coefficients

The correlation coefficient of a set of paired data measures how closely the regression line fits the data points.

A value close to $+1$ indicates a strong positive correlation.

A value close to -1 indicates a strong negative correlation.

The closer the value is to 0, the weaker is the correlation.

If two data pairs in a dataset are exactly the same, then they will be plotted on top of each other in a scatterplot, but the repetition is taken into account when the regression line and correlation coefficient are calculated.

In the next activity you are asked to find the correlation coefficient for the data on Robert Wadlow's age and height.

Activity 27 Finding the correlation coefficient for Mr Wadlow's data



Dataplotter

In Dataplotter, with the Scatterplot tab selected, use the drop-down lists to choose the datasets '# Wadlow age' and '# Wadlow ht' for the first and second columns, respectively. Find the correlation coefficient of these data by clicking the Regression box, if it is not already ticked. Comment on the value of the correlation coefficient.

Finding a regression line and examining the correlation coefficient are useful ways of analysing a set of paired data.

In the next activity you are asked to investigate the question: 'Is there a relationship between a mother's weight at the start and at the end of her pregnancy?' You have seen that after a question like this is posed, the next stage of the statistical modelling cycle is to collect some data. A dataset relevant to this question was introduced in Unit 4, and in the activity you are asked to use a regression line and correlation coefficient to analyse the data, and then to interpret the results that you obtain.

The relevant dataset from Unit 4 was referred to as the 'backache dataset', and you might remember that some data values were missing or clearly wrong and had to be removed from the dataset. When you plot a scatterplot using such data, you have to remember to remove data values

that are not paired. Dataplotter contains adjusted versions of some of the columns from the backache dataset, with these data values removed. They are indicated with the prefix 'SP' for 'scatterplot'. Before plotting any scatterplot, you should always check that both datasets contain the same number of values and that the values are paired correctly.



Dataplotter

Activity 28 *Investigating correlation in the backache dataset*

Use Dataplotter, with the Scatterplot tab selected.

- Obtain a scatterplot of the mothers' weights at the end of their pregnancies against their weights at the start, by choosing the datasets '# SP Weight start' and '# SP Weight end' for the first and second columns, respectively. Make sure that the Regression box is ticked.
- Write down the equation of the regression line, and the correlation coefficient.
- Use the equation of the regression line to predict the weight of a mother at the end of her pregnancy if her weight at the start was 55 kg. How reliable do you think your prediction is?

In the next activity you are asked to use the backache dataset from Unit 4 to investigate the relationship between the weight gain of mothers during their pregnancies, and the weights of their babies.



Dataplotter

Activity 29 *Investigating correlation in the backache dataset again*

Use Dataplotter, with the Scatterplot tab selected.

- Obtain a scatterplot of the babies' weights against the mothers' weight gains, by choosing the datasets '# SP Weight gain' and '# SP Weight baby' for the first and second columns, respectively. Make sure that the Regression box is ticked.
- What is the correlation coefficient? Describe the correlation between the mothers' weight gains and their babies' weights.

In the next activity you are asked to add a rogue measurement – an outlier – to the dataset investigated in the last activity, to investigate the effect that this has on the correlation coefficient.



Dataplotter

Activity 30 *Investigating the effect of an outlier in the data*

Use Dataplotter, with the Scatterplot tab selected, and the datasets '# SP Weight gain' and '# SP Weight baby' selected for the first and second columns, respectively, as in Activity 29. Make sure that the Regression box is ticked.

Change the name of the datasets '# SP Weight gain' and '# SP Weight baby' to 'SP Wt gain new' and 'SP Wt baby new', or new names of your own choice. You can do this by clicking on the blue or green dataset name above each dataset (not the dataset name in the drop-down list).

In Dataplotter you cannot directly change a preloaded dataset. To edit a preloaded dataset you first need to change its name.

- (a) Suppose that errors were made in recording one mother's data values, so that the dataset included a mother's weight gain recorded as 75 kg and her baby's weight recorded as 25 kg. Add this data pair to the data in the columns. What is the correlation coefficient of the new dataset, and how does it compare to the correlation coefficient of the original dataset, which you found in Activity 29?
- (b) Would it be better to use the regression line found in part (a), instead of the one in Activity 29, to predict a baby's weight from the mother's weight? Explain your answer.

When you use a regression line to make a prediction, you should not assume that if the correlation coefficient is close to 1 or -1 then the prediction is likely to be reliable. It is important to check the scatterplot as well.

You saw a reason for this in Activity 30: an outlier in the data can cause the correlation coefficient to have a value misleadingly close to 1 or -1 . You can see roughly why this happens if you compare the two scatterplots in the solutions to Activities 29 and 30. The data points are the same in each case except that an outlier has been added to the second plot. You can see that in the second plot the distances of the points from the line are much smaller in comparison to the overall range of the data, and this causes the improvement in the correlation coefficient.

Another reason why you need to check scatterplots as well as correlation coefficients is that there may be a clear relationship between two quantities that is not linear. For example, consider the data points in Figure 36. The regression line, which is shown, is a poor fit, and the correlation coefficient is also fairly poor, at 0.77. However, the data points seem to lie on a smooth curve, so it is not appropriate to model these data by a straight line – a curve would be a much better model.

You can learn about using curves to model paired data if you go on to take modules on statistics.

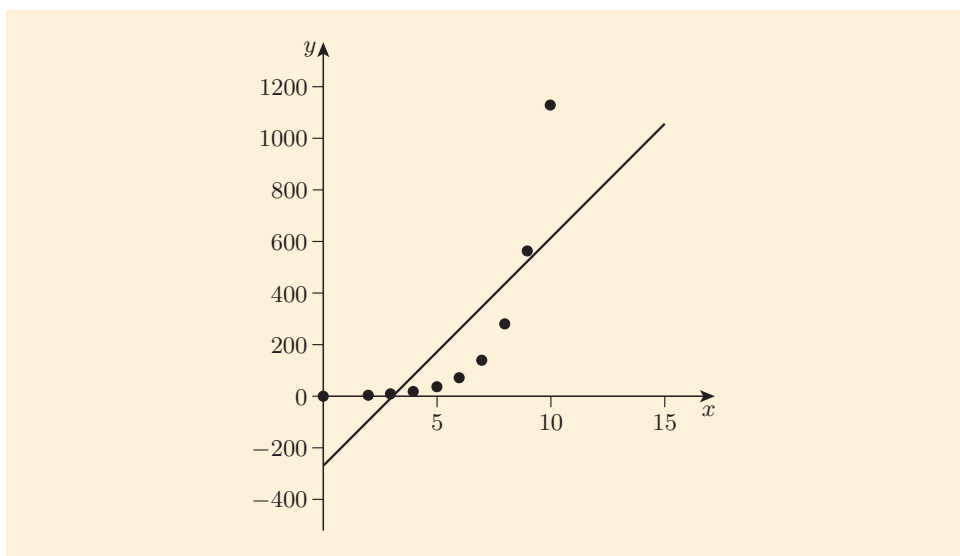


Figure 36 Data points that should not be modelled by a straight line

Correlation and causation

If there appears to be a strong correlation between two quantities, it does not necessarily mean that one has caused the other.

For example, suppose that data have been collected each year on the number of students in a particular town who are studying at least one level 1 mathematics module, and also on the number of burglaries committed in the same town. Suppose that these data give the scatterplot in Figure 37(a). Each data point corresponds to the figures for a particular year.

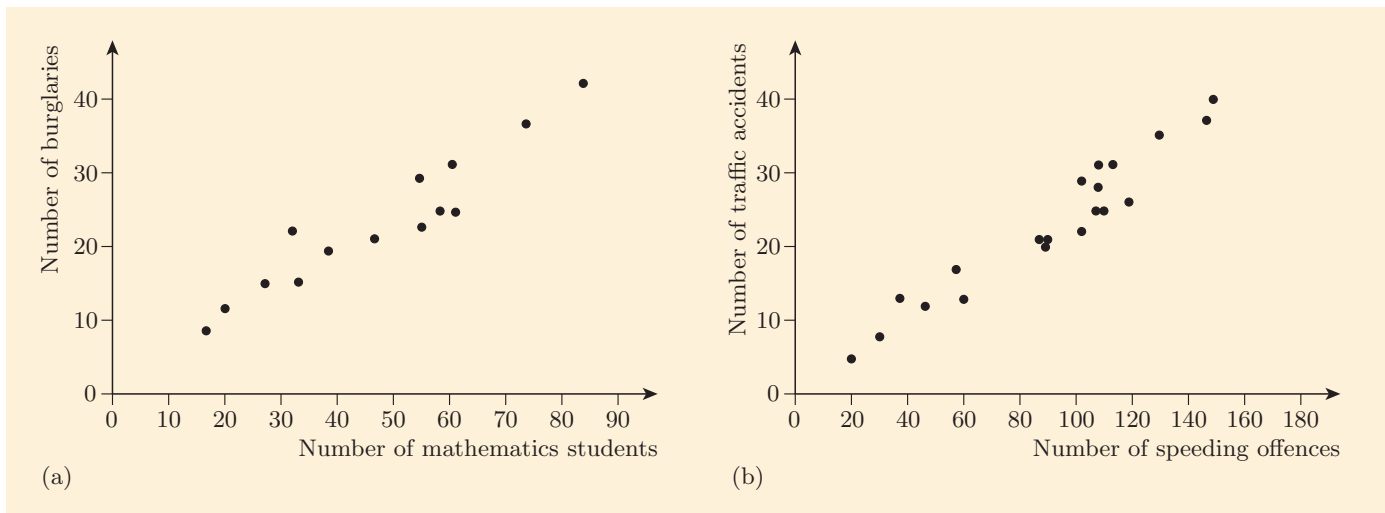


Figure 37 Two scatterplots for a fictitious town: (a) the annual number of burglaries plotted against the annual number of students studying level 1 mathematics modules; (b) the annual number of road accidents plotted against the annual number of recorded speeding offences

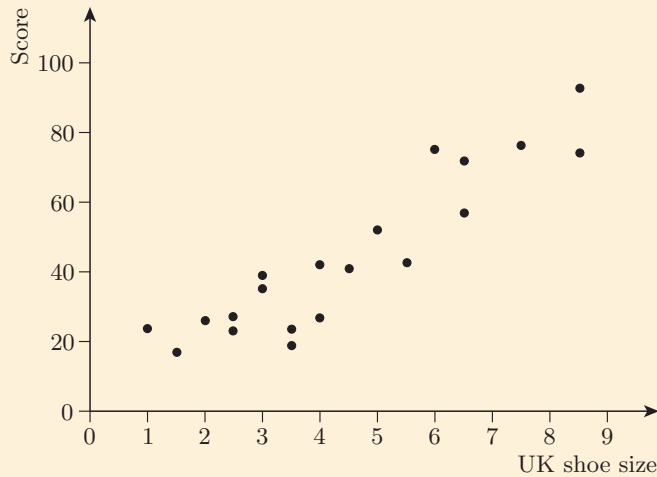
The graph in Figure 37(a) shows a strong correlation between the two quantities – the more mathematics students there are, the more burglaries are committed. However, you would not expect that the mathematics students are causing the burglaries! A likely explanation for the correlation is that both the number of students and the number of burglaries depend on some third quantity, such as the number of new people who are moving into the area.

Now look at the scatterplot in Figure 37(b), which shows the number of recorded speeding offences and the number of road accidents in the same town each year for several years. Again, there appears to be a strong correlation, and you might be tempted to conclude that speeding causes traffic accidents. However, just as before, the correlation does not prove a causal relationship. Again both quantities could depend on a third quantity: for example there might have been a varying number of newly-qualified drivers, and it could be that newly-qualified drivers are more likely both to exceed the speed limit and to cause road accidents. On the other hand, it *might* be true that speeding causes road accidents, but further evidence would be needed to establish this effect conclusively.

Activity 31 Comparing mathematical ability and shoe size

The scatterplot on the next page shows the scores achieved by some children in a mathematics test plotted against their shoe sizes.

- Describe the correlation between shoe size and mathematics score for these children.
- Can you conclude that the size of a child's feet determines his or her mathematical achievement?
- Give another possible explanation for the correlation observed.



Correlations between two quantities are frequently mentioned in the media, and it is often assumed that a correlation means that one quantity has caused the other. Whenever you hear of a correlation like this, you should ask yourself whether there might be any third quantity that could explain the correlation.

For example, it has been reported that children who were breastfed for longer as babies tend to have higher IQs (intelligence quotients). This correlation does not necessarily mean that breastfeeding improves a child's IQ. It could be that the kind of mother who breastfeeds for longer tends also to be the kind of mother who does other things that might improve her child's IQ, such as spending more time reading to and playing with her child.

In 1999 the journal *Nature* published a research paper that showed that babies who slept with the light on were more likely to develop myopia (nearsightedness) later in life than babies who slept in the dark. The authors concluded that sleeping with the light on can cause myopia. However, later research showed that sleeping in the light or dark had nothing to do with developing myopia, and the following alternative explanation of the correlation was suggested. Children with myopic parents are more likely to develop myopia themselves. And if you are a myopic parent, you are more likely to leave a light on in your child's room, so that you can find your way in and out!

In this section you have seen how to find the best straight line to model a set of paired data, and how to use the correlation coefficient to measure how well the line fits the data. Many situations can be modelled by straight lines in this way, particularly if the range of the dataset is small. However some data sets are modelled better by curved graphs, as you will see in Units 10 and 13.

5 Reviewing straight-line graphs

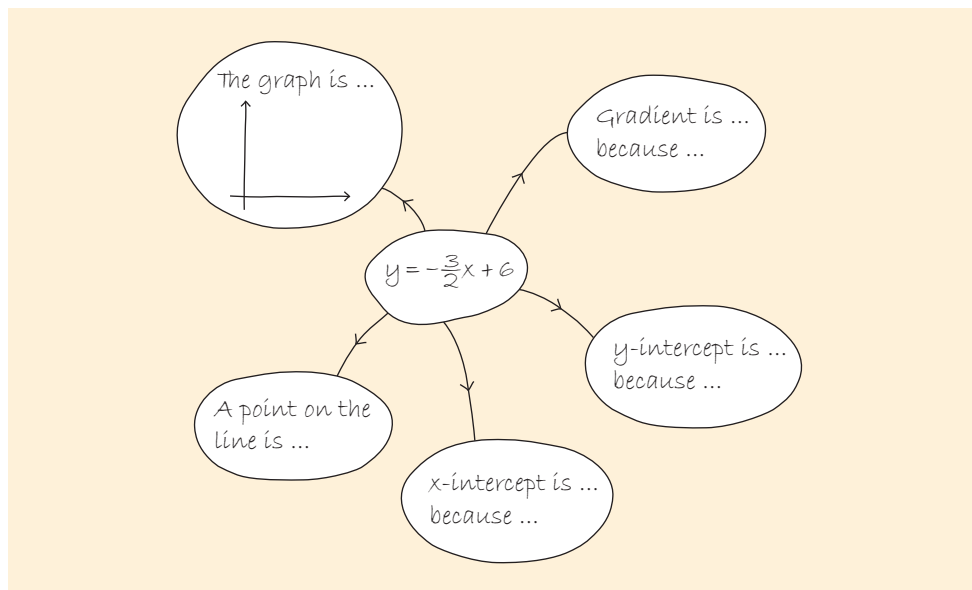
Graphs of straight lines, their equations and models based on them are used frequently in many different contexts, and you will work with them later in the module, particularly in Unit 7. So it is important that you are confident with the ideas and techniques covered in this unit.

Some people find that a helpful way of reviewing their progress on a topic is to draw a diagram showing the main ideas and the connections between them. Writing down the key ideas and thinking about how to get from one to another can strengthen your overall understanding, help you to remember the ideas more easily and provide you with a revision guide that may be useful when you need to look something up. You can include as much or as little detail in your diagram as you like. For instance, you can include examples, or use different colours for different aspects.

The next activity invites you to complete two such diagrams for some of the ideas in this unit.

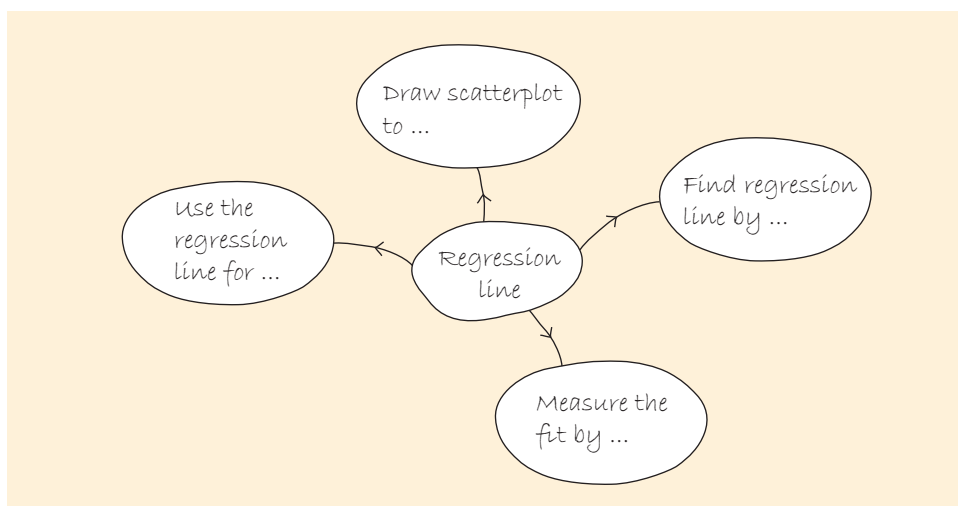
Activity 32 Reviewing straight-line graphs

- (a) Fill in the blanks in the incomplete diagram below, starting with the equation $y = -\frac{3}{2}x + 6$ in the middle.



- (b) Draw a diagram to summarise the ideas in Section 4. Decide what seems to be the central idea of this section – for example, regression lines – and put this idea in the centre of your diagram. Then add other ideas to this central idea as illustrated below.

What other ideas and information could you add? Use other colours or examples if you think it would help you.



Learning checklist

After studying this unit, you should be able to:

- plot a graph by constructing a table of values for a formula and plotting points
- calculate the gradient of a straight line on a graph and interpret gradient in practical situations
- find the intercepts of a straight line on a graph and interpret them in practical situations
- recognise direct proportion relationships and their graphs, and find a formula for a direct proportion relationship
- find the intercepts and gradient of a straight line from its equation, and find the equation of a straight line from its gradient and y -intercept
- use the equation of a straight line to draw that line on a graph, either by using the y -intercept and gradient from the equation, or by plotting two points
- find the equation of a straight line from its gradient and a point on the line, or from two points on the line
- use the Scatterplot page of Dataplotter to find the regression line and correlation coefficient for a set of paired data, and interpret the results
- interpolate values sensibly and be aware of the dangers of extrapolation.

Solutions and comments on Activities

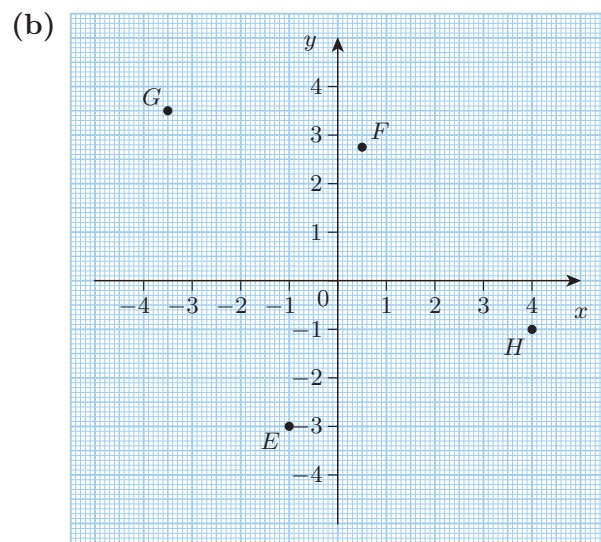
Activity 1

(a) A has coordinates $(1.5, 1)$.

B has coordinates $(-2, 3)$.

C has coordinates $(-3.5, -2)$.

D has coordinates $(2.5, -3.5)$.



Activity 2

(a) Substituting $x = 6$ into the equation gives

$$y = \frac{1}{2} \times 6 + 3 = 3 + 3 = 6.$$

So the point $(6, 5)$ does not lie on the line.

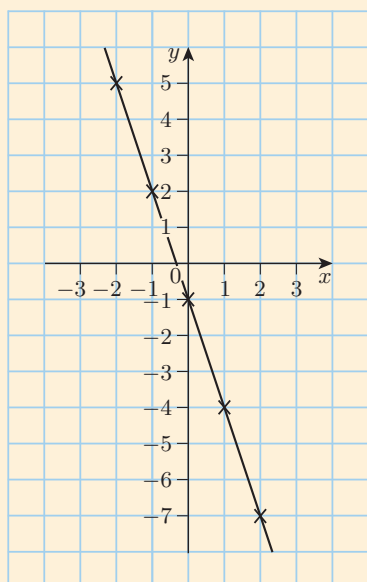
(b) Substituting $x = -5$ into the equation gives

$$y = \frac{1}{2} \times (-5) + 3 = -2.5 + 3 = 0.5.$$

So the point $(-5, 0.5)$ lies on the line.

Activity 3

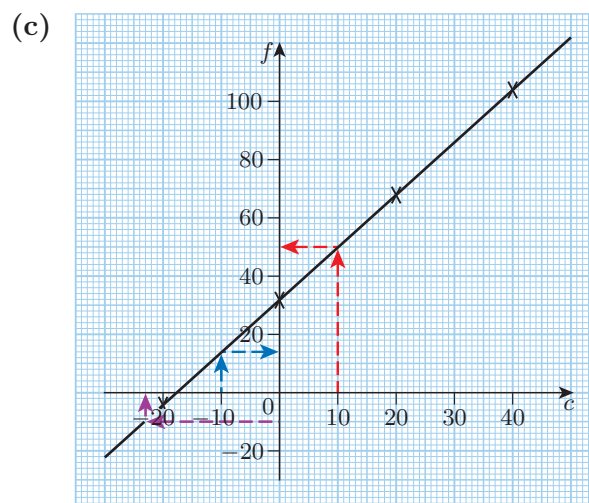
x	-2	-1	0	1	2
y	5	2	-1	-4	-7



Activity 4

(a) The independent variable is c , since f is the subject of the equation.

c	-20	0	20	40
f	-4	32	68	104



(d) (i) A temperature of 10°C is 50°F . (This conversion is shown in red on the graph.)

(ii) A temperature of -10°C is 14°F . (This conversion is shown in blue on the graph.)

(iii) A temperature of -10°F is approximately -23°C . (This conversion is shown in purple on the graph.)

(iv) A temperature of 0°F is approximately -18°C . (This is the value where the graph crosses the c -axis.)

Activity 5

(a) The line in Figure 3(b) predicts that if the greengrocer charges $\pounds 3.50$ per kg for tomatoes, then she should expect to sell about 770 kg.

(b) Similarly, if the greengrocer charges $\pounds 4.75$ per kg for tomatoes, then she should expect to sell about 600 kg.

Activity 6

(a) (i) The run is $3 - 1.5 = 1.5$. The rise is $2 - 6 = -4$. So the gradient is

$$\frac{\text{rise}}{\text{run}} = \frac{-4}{1.5} = \frac{-40}{15} = -\frac{8}{3}.$$

(Check: The line slopes down, so the gradient should be negative.)

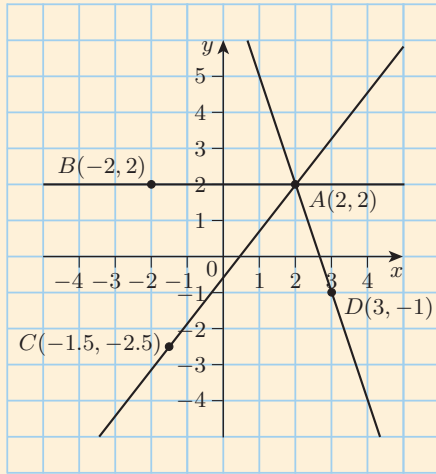
- (ii) The run is $1 - (-1.5) = 1 + 1.5 = 2.5$.
The rise is $2 - (-2) = 4$.

So the gradient is

$$\frac{\text{rise}}{\text{run}} = \frac{4}{2.5} = 1.6.$$

(Check: The line slopes up, so the gradient should be positive.)

(b)



- (i) The gradient of the line that passes through A and B is

$$\frac{\text{rise}}{\text{run}} = \frac{2 - 2}{2 - (-2)} = \frac{0}{4} = 0.$$

- (ii) The gradient of the line that passes through A and C is

$$\frac{\text{rise}}{\text{run}} = \frac{2 - (-2.5)}{2 - (-1.5)} = \frac{4.5}{3.5} = \frac{45}{35} = \frac{9}{7} = 1\frac{2}{7}.$$

- (iii) The gradient of the line that passes through A and D is

$$\frac{\text{rise}}{\text{run}} = \frac{-1 - 2}{3 - 2} = \frac{-3}{1} = -3.$$

- (c) The rise between any two points on a horizontal line is zero. Because the gradient is the rise divided by the run, it follows that the gradient of a horizontal line is zero.

Activity 7

- (a) This line slopes up, so it has a positive gradient, and the angle that it makes with the x -axis is greater than 45° , so its gradient is greater than 1.

- (b) This line slopes down, so it has a negative gradient, and the angle that it makes with the x -axis seems to be about 45° , so its gradient is about -1 .

- (c) This line slopes up, so it has a positive gradient, and the angle that it makes with the x -axis is less than 45° , so its gradient is less than 1.

That is, its gradient is between 0 and 1.

- (d) This line slopes down, so it has a negative gradient, and the angle that it makes with the x -axis is greater than 45° , so its gradient is greater than 1 in size. That is, its gradient is less than -1 .

Activity 8

- (a) A is $(-0.6, 2)$; B is $(0.4, 1)$; C is $(-0.4, -2)$; D is $(0.2, 5)$.

- (b) (i) The gradient of the line through A and B is

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{1 - 2}{0.4 - (-0.6)} = \frac{-1}{1} = -1.$$

- (ii) The gradient of the line through A and D is

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{5 - 2}{0.2 - (-0.6)} = \frac{3}{0.8} = 3.75.$$

- (iii) The gradient of the line through B and C is

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{-2 - 1}{-0.4 - 0.4} = \frac{-3}{-0.8} = 3.75.$$

- (c) The gradients of these two lines are the same.

Activity 9

- (a) The points $(0, 0)$ and $(30, 600)$ lie on the line, so its gradient is

$$\frac{600 - 0}{30 - 0} = \frac{600}{30} = 20 \text{ km/l.}$$

The gradient measures the number of kilometres travelled per litre of fuel. In other words, the gradient is the rate of fuel consumption.

- (b) The points $(0, 0)$ and $(2, 80)$ lie on the line, so its gradient is

$$\frac{80 - 0}{2 - 0} = \frac{80}{2} = 40 \text{ km/h.}$$

The gradient measures the number of kilometres travelled per hour. In other words, the gradient is the speed.

Activity 10

- (a) The vertical intercept is 25 km.

This is the distance of the walker from the finish line at the start of the walk. In other words, it is the length of the walk.

The horizontal intercept is 5 hours.

This is the time when the distance of the walker from the finish line is zero. In other words, it is the time taken for the walker to complete the walk.

- (b) The vertical intercept is £210.

This is the cost of hiring the venue for no people. (The total cost is this cost plus an amount that depends on the number of people using the venue.)

Activity 11

(c) The line $y = 0.5x$ makes the smallest angle with the x -axis.

Activity 12

(a) The time and the distance are directly proportional. If you multiply or divide the time by any number, then the distance changes in the same way.

(b) The number of painters and the time taken are *not* directly proportional. The more painters there are, the less time the job takes. (In theory, anyway!)

(c) The number of pounds exchanged and the number of euros received are *not* directly proportional. If you double the number of pounds, then the number of euros that you receive is not doubled, because of the transaction fee. For example, if the exchange rate is €1.20 to the pound and you exchange £100, then you receive

$$100 \times \text{€}1.20 - \text{€}10 = \text{€}110,$$

whereas if you exchange £200, then you receive

$$200 \times \text{€}1.20 - \text{€}10 = \text{€}230,$$

which is not double €110.

(d) The number of songs and the total cost are directly proportional.

(e) The number of hours worked and the gross pay are directly proportional.

Activity 13

(a) Since d is directly proportional to c , the formula expressing their relationship is of the form

$$d = kc,$$

where k is a constant.

Also, $d = 400$ when $c = 20$. Substituting these values into the equation gives

$$400 = 20k.$$

We now solve this equation to find the value of k .

$$\text{Divide by 20: } \frac{400}{20} = k$$

$$\text{Simplify: } 20 = k$$

So the required formula is

$$d = 20c.$$

(b) Substituting $c = 35$ into the formula found in part (a) gives

$$d = 20 \times 35 = 700.$$

So the amount of drug needed is 700 mg.

Activity 14

(a) (i) This graph is a straight line but it does not pass through the origin, so it does not represent a direct proportion relationship.

(ii) This graph is not a straight line so it does not represent a direct proportion relationship.

(iii) This graph is a straight line through the origin so it represents a direct proportion relationship.

(b) Temperature in degrees Fahrenheit is not directly proportional to temperature in degrees Celsius, because the graph of this relationship does not pass through the origin.

Activity 15

(b) The line through the origin with gradient 2 is shown in Figure 21 on page 85. If you have not obtained this graph, then check that you have chosen the correct options in Graphplotter.

(c) The line moves 1 unit up the y -axis. The new graph has y -intercept 1.

(d) You should find that changing the value of c moves the graph up or down the y -axis. The y -intercept is always the value of c .

(e) The lines all have the same gradient (that is, they are all parallel to each other), but they have different y -intercepts.

(f) Changing the value of m changes the gradient, and changing the value of c changes the y -intercept.

Activity 16

(a) (i) The coefficient of x is 2, so the gradient is 2. The constant is -1 , so the y -intercept is -1 .

(ii) The coefficient of x is -3 , so the gradient is -3 . The constant is 4, so the y -intercept is 4.

(iii) The coefficient of x is $\frac{1}{5}$, so the gradient is $\frac{1}{5}$. The constant is $-\frac{2}{5}$, so the y -intercept is $-\frac{2}{5}$.

(b) (i) The equation is $y = 4x - 10$.

(ii) The equation is $y = -x + 5$.

(iii) The equation is $y = 0x + 3$; that is, $y = 3$.

(c) (i) The equation is $y = 4x + 2$.

(ii) The equation is $y = -x + 2$.

(iii) The equation is $y = 2$.

Activity 17

(a) Putting $y = 0$ gives
 $2x - 1 = 0$.

We now solve this equation.

$$\text{Add 1: } 2x = 1$$

$$\text{Divide by 2: } x = \frac{1}{2}$$

Hence the x -intercept is $\frac{1}{2}$.

- (b) Putting $y = 0$ gives
 $-3x + 4 = 0$.

We now solve this equation.

Add $3x$: $4 = 3x$

Divide by 3: $\frac{4}{3} = x$

Hence the x -intercept is $\frac{4}{3}$.

- (c) Putting $y = 0$ gives

$$\frac{x}{5} - \frac{2}{5} = 0.$$

We now solve this equation.

Multiply by 5: $x - 2 = 0$

Add 2: $x = 2$

Hence the x -intercept is 2.

Activity 18

- (a) The equations of the lines are as follows.

(i) $y = 1$

(ii) $x = -3$

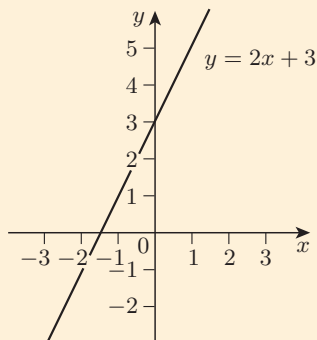
(iii) $y = -4.5$

(iv) $x = 2.3$

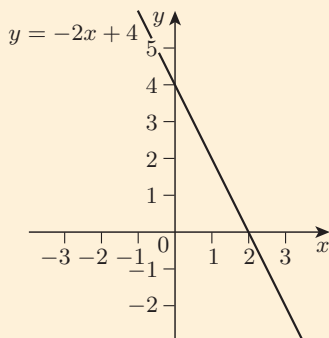
- (b) The x -axis has equation $y = 0$, and the y -axis has equation $x = 0$.

Activity 19

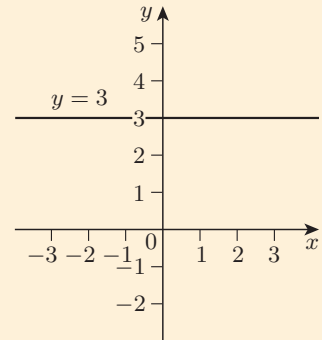
- (a)



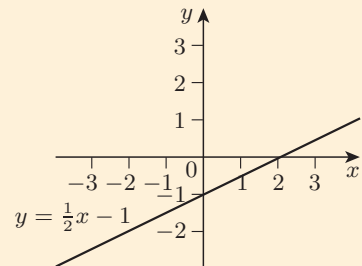
- (b)



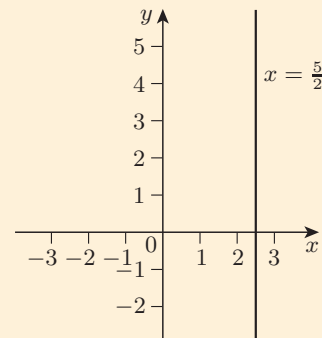
- (c)



- (d)



- (e)



Activity 20

- (a) The line passes through the points $(0, 1)$ and $(1, -2)$. So the gradient is

$$\frac{-2 - 1}{1 - 0} = \frac{-3}{1} = -3.$$

The y -intercept is 1.

Hence the equation is $y = -3x + 1$.

- (b) The line passes through the points $(-0.5, -0.8)$ and $(0, 0.3)$. So the gradient is

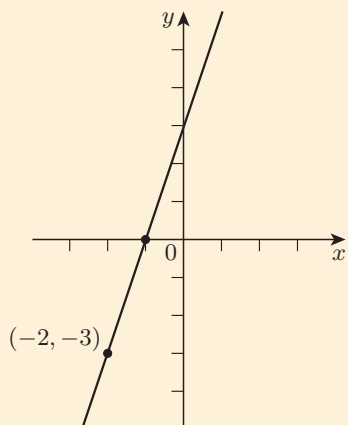
$$\frac{0.3 - (-0.8)}{0 - (-0.5)} = \frac{1.1}{0.5} = 2.2.$$

The vertical intercept is 0.3.

Hence the equation is $s = 2.2t + 0.3$.

Activity 21

A sketch of the line is as follows.



The gradient of the line is 3, so the equation is $y = 3x + c$, where c is a constant.

Also, the line passes through the point $(-2, -3)$. Substituting these coordinates into the equation gives

$$-3 = 3 \times (-2) + c; \quad \text{that is,} \quad -3 = -6 + c.$$

Adding 6 to both sides gives

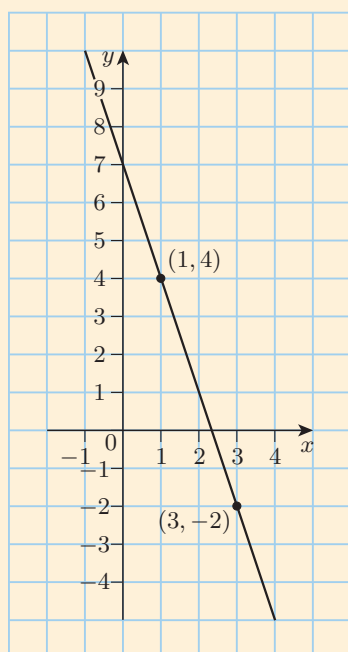
$$3 = c.$$

So the equation of the line is $y = 3x + 3$.

(Check: This is the equation of a line with y -intercept 3, and the y -intercept on the sketch does appear to be 3.)

Activity 22

(a)



(b) The gradient of the line is

$$\frac{-2 - 4}{3 - 1} = \frac{-6}{2} = -3.$$

So the equation of the line is $y = -3x + c$, where c is a constant.

Also, the line passes through the point $(1, 4)$.

Substituting these coordinates into the equation gives

$$4 = -3 \times 1 + c; \quad \text{that is,} \quad 4 = -3 + c.$$

Adding 3 to both sides gives

$$7 = c.$$

So the equation of the line is

$$y = -3x + 7.$$

(Check: Substituting $x = 3$ into the equation gives

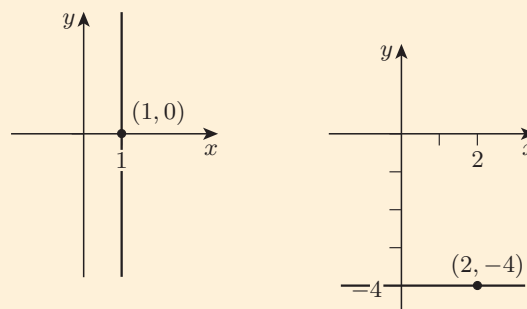
$$y = -3 \times 3 + 7 = -9 + 7 = -2,$$

so the coordinates $(3, -2)$ satisfy the equation.)

Activity 23

(a) A sketch of the line is shown in the left-hand figure below. The equation is $x = a$, where a is a constant. Since the point $(1, 0)$ satisfies the equation, the equation is $x = 1$.

(b) A sketch of the line is shown in the right-hand figure below. The equation is $y = a$, where a is a constant. Since the point $(2, -4)$ satisfies the equation, the equation is $y = -4$.

**Activity 24**

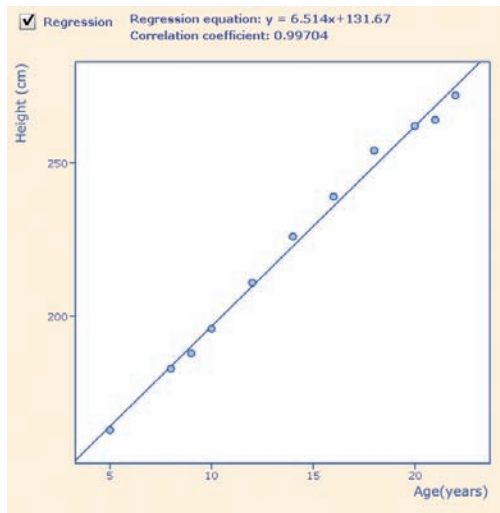
(a) You should obtain a scatterplot similar to Figure 28 on page 104, as shown on the next page.

(b) The equation of the regression line is

$$y = 6.5140x + 131.67,$$

where x and y represent Robert Wadlow's age in years and height in centimetres, respectively.

(Here the coefficient of x and the constant term are given to five significant figures. This is the default setting of the Scatterplot page of Dataplotter.)



Activity 25

(a) Substituting $x = 24$ into the equation of the line gives

$$y = 6.5140 \times 24 + 131.67 = 288.006.$$

So Mr Wadlow's height at age 24 is predicted to be about 288 cm.

(b) Moving the cursor over the Dataplotter graph to find coordinates gives a similar answer to that found in part (a).

(c) The age of 24 is outside the range of the data used to produce the model. Mr Wadlow's growth may not have continued at the same rate, so the estimate provided by the regression line for his height at age 24 may not be reliable. (In fact, sadly Mr Wadlow died when he was 22, following an infection in a blister caused by a leg brace.)

Activity 26

(b) The value of the correlation coefficient is 1.

(c) The regression line and correlation coefficient usually change each time a new point is plotted. Once there is a data point not on the regression line, the value of the correlation coefficient is not 1, but some value between -1 and 1 .

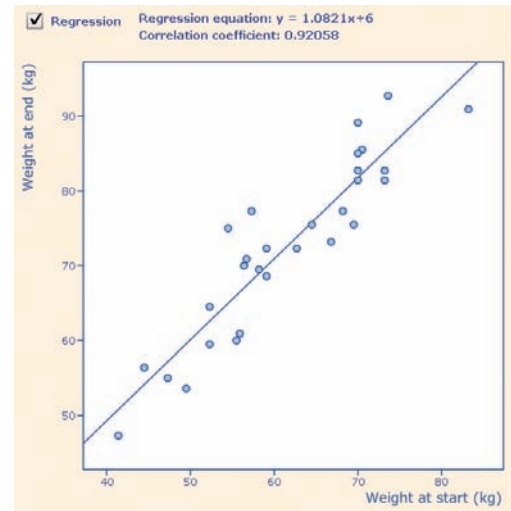
(d) This time the points on the line have a correlation coefficient of -1 . Again, the regression line and correlation coefficient usually change each time a new point is plotted. Once there is a data point not on the regression line, the value of the correlation coefficient is not -1 , but again some value between -1 and 1 .

Activity 27

The correlation coefficient for the data on Mr Wadlow's age and height is 0.99704 (to 5 s.f.). This value is close to 1, so it indicates that the regression line is a good fit to the data points.

Activity 28

(a) The scatterplot is shown below.



(b) The equation of the regression line is

$$y = 1.0821x + 6.0000,$$

and the correlation coefficient is 0.92058 (both to five significant figures).

(c) Substituting $x = 55$ into the equation of the regression line gives

$$y = 1.0821 \times 55 + 6.0000 = 65.5155.$$

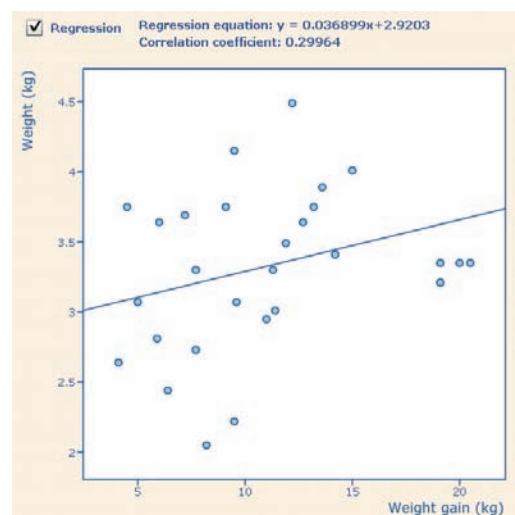
So the model predicts that a mother weighing 55 kg at the start of her pregnancy will weigh about 66 kg at the end.

The correlation coefficient is only moderately high, at about 0.92, so the prediction is unlikely to be very accurate, though it does give some indication of the expected weight.

(However, predictions made using a regression line may not be reliable for individuals.)

Activity 29

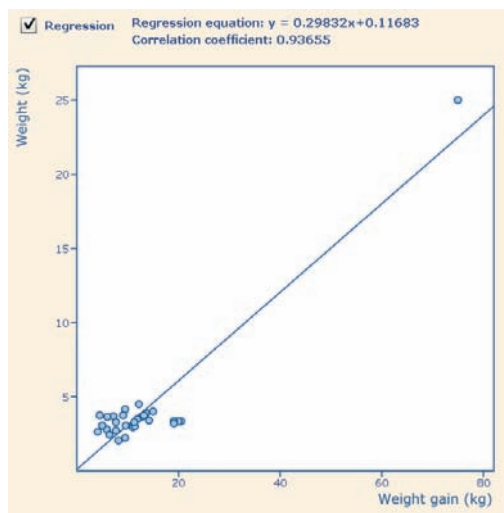
(a) The scatterplot is shown below.



(b) The correlation coefficient is only 0.29964 (to 5 s.f.), that is, about 0.30, which is fairly close to zero, so there appears to be little correlation between the mothers' weight gains and their babies' weights. This is confirmed by the scatterplot, which shows a large amount of scatter.

Activity 30

(a) The new scatterplot is shown below.



The new correlation coefficient is 0.93655 (to 5 s.f.), that is, about 0.94, which is significantly higher than the correlation coefficient for the original dataset, which is about 0.30.

(b) Despite the higher correlation coefficient in part (a), it would be better to use the regression line in Activity 29 to make a prediction, since the regression line in part (a) is based on the same data together with a misleading data pair. However, since there is little correlation shown by the data in Activity 29, it would be unwise to use either regression line to make a prediction.

(Some explanation of the effect of adding the erroneous data is given after the activity.)

Activity 31

(a) There appears to be a strong positive correlation between shoe size and mathematics score.

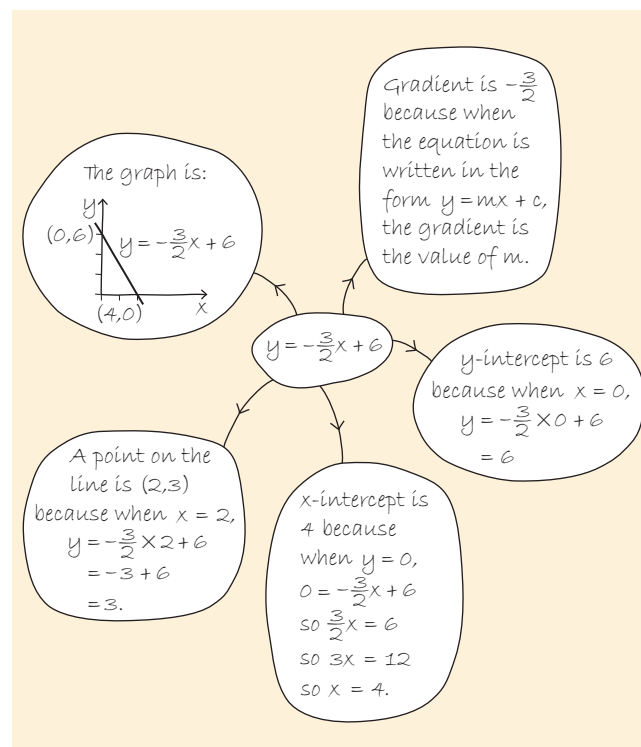
(b) It is not reasonable to conclude that foot size determines mathematical achievement. Correlation does not imply causation.

(c) A likely explanation for the correlation is that both foot size and mathematical achievement increase as a child gets older.

Activity 32

(a) An example of a completed diagram is shown below, but yours will probably be different to this

one – for example, you might have chosen a different point on the line, or explained your answers in slightly different ways.



(b) One suggestion is below – notice how extra information has been added. You may wish to add other details, such as an example of how to substitute a value into the equation of a regression line, or further information on using the Scatterplot page in Dataplotter. The aim of a diagram like this is to include the main points and anything else that will help you as a revision aid.

